Flexibook

Color Atlas of Genetics

Eberhard Passarge

Fifth Edition

basic sciences

Thieme

At a Glance

Introduction	1
Fundamentals	25
Prologue	26
Molecular Basis of Genetics	36
Analysis of DNA	58
Variability of DNA	72
Processing of DNA	80
Eukaryotic Cells	86
Formal Genetics	104
Chromosomes	132
Regulation of Gene Function	162
Epigenetic Modifications	178
Genetic Signal Pathways	186
Genes in Embryonic Development	196
Genomics	205
Genetics in Medicine	235
Genetic Classification of Diseases	236
Imbalanced Homeostasis	260
Metabolic Disorders	276
Immune System	296
Origins of Cancer	312
Impaired Cell and Tissue Structure	334
Hemoglobin Disorders	350
Sex Determination and Differentiation	362
Sensory Perception	370
Chromosomal Aberrations	382
A Brief Guide to Genetic Diagnosis	388
Morbid Anatomy of the Human Genome	392
Chromosomal Locations—Alphabetical List	398
Appendix	403
Glossary	421
Index	447

Bernhard Horsthemke
(“Benno”)

in appreciation of thirty years
of successful work together

Color Atlas of Genetics

Eberhard Passarge, MD
Professor of Human Genetics
Emeritus Director
Institute of Human Genetics
University Hospital Essen
Essen, Germany

Fifth edition, revised and updated
With 186 color plates prepared
by Jürgen Wirth

Thieme
Stuttgart · New York

Library of Congress Cataloging-in-Publication Data is available from the publisher.

Previous and foreign editions of this book:
1st German edition 1994
1st English edition 1995
1st French edition 1995
1st Japanese edition 1996
1st Chinese edition 1998
1st Italian edition 1999
1st Turkish edition 2000
2nd English edition 2001
2nd French edition 2003
1st Portuguese edition 2004
1st Spanish edition 2004
2nd German edition 2004
1st Polish edition 2004
1st Greek edition 2005
1st Arabic edition 2006
3rd English edition 2007
4th English edition 2013
3rd German edition 2008
3rd French edition 2008
2nd Turkish edition 2009
2nd Japanese edition 2009
2nd Spanish edition 2010
2nd Portuguese edition 2010
3rd Turkish edition 2017

Note: The English and the German editions are written by the author and are not translations.

© 2018 Georg Thieme Verlag KG,
Rüdigerstrasse 14, 70469 Stuttgart, Germany
http://www.thieme.de
Thieme Medical Publishers, Inc., 333 Seventh Avenue,
New York, NY 10001, USA
http://www.thieme.com

Color plates prepared by Jürgen Wirth, Professor of Visual Communication, Dreieich, Germany

Cover design: Thieme Publishing Group
Typesetting by Thomson Digital, India
Printed in China by Everbest Printing Ltd., Hong Kong

ISBN 978-3-13-241440-2
eISBN 978-3-13-241441-9

Important note: Medicine is an ever-changing science undergoing continual development. Research and clinical experience are continually expanding our knowledge, in particular our knowledge of proper treatment and drug therapy. Insofar as this book mentions any dosage or application, readers may rest assured that the authors, editors, and publishers have made every effort to ensure that such references are in accordance with **the state of knowledge at the time of production of the book.**

Nevertheless, this does not involve, imply, or express any guarantee or responsibility on the part of the publishers in respect to any dosage instructions and forms of applications stated in the book. **Every user is requested to examine carefully** the manufacturers' leaflets accompanying each drug and to check, if necessary in consultation with a physician or specialist, whether the dosage schedules mentioned therein or the contraindications stated by the manufacturers differ from the statements made in the present book. Such examination is particularly important with drugs that are either rarely used or have been newly released on the market. Every dosage schedule or every form of application used is entirely at the user's own risk and responsibility. The authors and publishers request every user to report to the publishers any discrepancies or inaccuracies noticed. If errors in this work are found after publication, errata will be posted at www.thieme.com on the product description page.

Some of the product names, patents, and registered designs referred to in this book are in fact registered trademarks or proprietary names even though specific reference to this fact is not always made in the text. Therefore, the appearance of a name without designation as proprietary is not to be construed as a representation by the publisher that it is in the public domain.

This book, including all parts thereof, is legally protected by copyright. Any use, exploitation, or commercialization outside the narrow limits set by copyright legislation, without the publisher's consent, is illegal and liable to prosecution. This applies in particular to photostat reproduction, copying, mimeographing, preparation of microfilms, and electronic data processing and storage.

Preface

As its preceding editions of 1995, 2001, 2007, and 2013, this small book provides an overview of the field of genetics, including selected aspects of genomics. It is based on a visual approach using 186 color plates designed by the author and graphically prepared for print by Jürgen Wirth, a professor of Visual Communication. Each plate corresponds to a small chapter illustrating a concept and related facts. An explanatory text accompanies each plate on its opposite page.

The subjects of the plates have been chosen based on their importance as fundamentals and their role in the understanding the genetic bases of inherited diseases. Owing to the limited space, individual diseases are not described in detail, but references are provided for further information. In addition, the corresponding Online Mendelian Inheritance of Man (OMIM) number is provided for each disease mentioned. The OMIM is a catalog of human genes and phenotypes introduced by Victor A. McKusick in 1966. It is freely available online as Online Mendelian Inheritance of Man (OMIM: www.ncbi.nlm.nih.gov/omim). It provides all genetically relevant information about the known genetic diseases (see p. 392).

This book maintains the general structure of the previous editions: Part I addresses Fundamentals; Part II, Genomics; and Part III, Genetics in Medicine. Part III illustrates the role of genetic and genomic principles underlying the causes of human diseases. From a genetic point of view a disease can be classified on the basis of its genetic causes (genotype) rather than its manifestations (phenotype), as is otherwise customary in medicine.

The book presents ancillary information in the Introduction. Genetics and genomics as viewed today are defined and some key developments of the past are traced. The Chronology specifically lists important discoveries in the history of genetics and genomics. The historical perspectives are a reminder that the platform of knowledge today rests on previous advances.

The Appendix provides tables with supplementary genetic data. The extensive Glossary defines genetic terms. For young readers naturally interested in the future, whenever possible and appropriate, I have included a historical perspective by referring to the first description of a discovery.

This fifth edition has been extensively rewritten, reorganized, and updated. Nineteen plates are entirely new or have new parts. New topics, represented by new plates, include overviews of human evolution, aging, the CRISPR-Cas principle, genetic signaling pathways, genomic disorders and genome-wide association studies, cancer genomes, laminopathies, chromatin disorders, cohesinoptathies, and other emerging topics. About the same number of plates have also been deleted because they are no longer needed. The fifth edition is slightly smaller than the fourth edition of 2013.

This book is written for two kinds of readers: for students of biology or medicine, as an introductory overview, and for their mentors, as a visual teaching aid. It will also help other interested individuals obtain selected information about current developments and achievements in this rapidly evolving field. The reader should keep in mind that each plate and its text represent an abstract rather than a treatise, with many related details necessarily omitted. Therefore, this book is meant to be a supplement to classic textbooks rather than a substitute.

The term *Atlas* for a book was introduced in 1594 by Gerard de Kremer (1512-1594), a Flemish mathematician and cartographer also known as Mercator. His book, with a collection of 107 double page geographic maps with the title *Atlas sive Cosmographicae Meditationes de Fabrica Mundi et Fabrica Figura,* was published in 1595, a year after his death. With Africa, Asia, the "New World", and the northern polar region represented by only one map each, it was the first world atlas. Mercator explains in his introduction that he derived the term from the mythic king, Atlas of Mauretania

because of his outstanding knowledge of astronomy. Earlier it was assumed that the term atlas referred to the titan, Atlas, of Greek mythology. When Mercator's atlas appeared, many geographic regions were not yet known and had remained unmapped in his collection. Establishing genetic maps is an activity not unlike mapping new, unknown territories 500 years ago. Genetic maps are a leitmotif in genetics and a recurrent theme in this book.

Throughout the book I have emphasized the role of the evolution of genes, genomes, and organisms in understanding genetics. As noted by the great geneticist Theodosius Dobzhansky, "Nothing in biology makes sense except in the light of evolution." Indeed, genetics and the science of evolution are closely related. Today one could say, "Nothing in evolution makes sense except in the light of genetics." As a single-author book, this book represents a personal view that has developed over a period of more than fifty years of active participation in the field.

I am deeply indebted to Professor Jürgen Wirth, Professor of Visual Communication at the Universities of Applied Sciences, Darmstadt and Schwäbisch Gmünd, Germany (during the period 1978 to 2005) for his most skillful work, which is a fundamental part of this book. I thank my wife, Mary Fetter Passarge, M.D., for her helpful suggestions. At Thieme Publishers, Stuttgart, I was guided and supported by Stephan Konnry, Andreas Schabert, Nidhi Chopra, Apoorva Goel, and others.

Eberhard Passarge

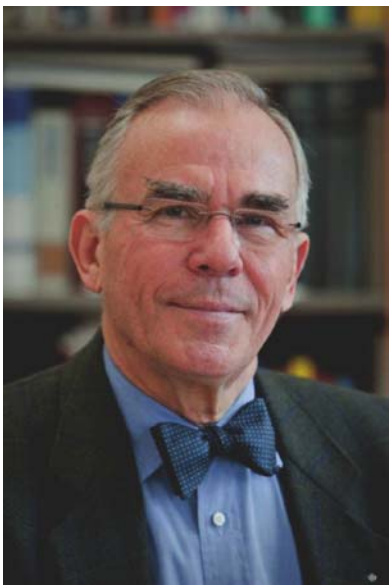
Acknowledgments

In preparing this fifth edition, I was supported by several colleagues who provided photographs and/or advice with information and text review. I am highly indebted to Gabriele Gillessen-Kaesbach and Frank Kaiser, Lübeck; Bernhard Horsthemke, Essen; Wolfram Kress, Würzburg; Stefan Mundlos, Berlin; Gesa Schwanitz, Bonn; Anne-Christin Teichmann, Sieglinde Siegert, and Andreas Scheduikat, Leipzig; and DagmarWieczorek, Düsseldorf.

For the previous editions, I would sincerely like to thank Mohammad Reza Ahmadian (Institute of Biochemistry and Molecular Biology II, University Hospital Düsseldorf, Germany); Beate Albrecht, Tea Berulava, and Bernhard Horsthemke (Institute for Human Genetics, Essen, Germany); Thomas Langmann (Pro Retina Professorship, Institute of Human Genetics, University of Regensburg, Germany); Maximilian Muenke (Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States); Heike Olbrich and Heymut Omran (Department of Pediatrics, University of Münster, Germany); and Martin Zenker (Institute of Human Genetics, University of Magdeburg, Germany). Their contributions are also acknowledged in the text accompanying the relevant plates.

Last but not least, the following colleagues deserve a mention for their generous contribution of valuable advice and additional information: Karin Buiting, Stephanie Gkalympoudis, Deniz Kanber, Dietmar Lohmann, Hermann-Josef Lüdecke, and Nicholas Wagner (Essen, Germany); Nicholas Katsanis (Duke University, Durham, North Carolina, United States); James R. Lupski (Houston, United States); Maximilian Muenke (Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States); Arne Pfeuffer (Helmholtz-Zentrum, Munich, Germany); and Friedrich Stock (Leipzig, Germany). A token of appreciation goes to the following colleagues for providing photographs or drafts for illustrations received previously for earlier editions, which are also used in the fourth edition: Alireza Baradaran (Mashhad, Iran, now Vancouver, Canada); Dirk Bootsma (Rotterdam, Netherlands); Laura Carrel (Hershey, Pennsylvania, United States); Aravinda Chakravarti and Richard I. Kelley (Baltimore, United States); Thomas Cremer (Munich, Germany); Robin Edison (National Institutes of Health, Bethesda, Maryland, United States); Evan E. Eichler (Seattle, United States); Wolfgang Engel (Göttingen, Germany); Reiner Johannisson (Lübeck, Germany); Nikolaus Konjetzko, Alma Küchler, Dietmar Lohmann, Axel Schneider, and Dagmar Wieczorek (Essen, Germany); Nicole McNeil and Thomas Ried (National Institutes of Health, Bethesda, Maryland, United States); Clemens Müller-Reible (Würzburg, Germany); Stefan Mundlos (Berlin, Germany); Helga Rehder (Marburg, Germany/Vienna, Austria); David L. Rimoin (Los Angeles, United States); Evelin Schröck (Dresden, Germany); Peter Steinbach (Ulm, Germany); Hans Hilger Ropers (Berlin, Germany); Sabine Uhrig and Michael Speicher (Graz, Austria); Michael Weis (Cleveland, Ohio, United States); and Eberhard Zrenner (Tübingen, Germany).

About the Author



Eberhard Passarge, MD, is a German human geneticist at the University Institute of Human Genetics at Essen, Germany. He graduated from the University of Freiburg, Germany in 1960 with an MD degree and received general medical training at the General Hospital Hamburg-Harburg, Germany (1961–1962) and at the Worcester Memorial Hospital, Worcester, Massachusetts, United States (1962–1963) with a stipend from the Ventnor Foundation. His postgraduate education was in pediatrics from the Cincinnati Children's Hospital Medical Center, Ohio, United States (1963–1966) with Josef Warkany and in human genetics from the Cornell Medical Center, New York, United States (1966–1968) with James German. After completing his training, he established a new division of Cytogenetics and Clinical Genetics at the Department of Human Genetics, University of Hamburg, Germany in 1968 and directed it until 1976, when he became the Professor of Human Genetics and the Founding Chairman of the Institute of Human Genetics, University of Essen, Germany, where he served until his retirement in 2001. He has remained active in the field of human genetics. From 2010 to 2014, he was the Interim Chairman at the Institute of Human Genetics at the University of Leipzig, Germany. Among his main scientific interests are the investigation of hereditary and congenital diseases and their application in genetic diagnosis and counseling. He is the author or coauthor of more than 250 scientific articles in international, peer-reviewed journals; the author of chapters in several international textbooks; and the author of three books on human and medical genetics. His experience in teaching human genetics is reflected in the *Color Atlas of Genetics*. He has served on the editorial board of several international human genetic journals. He has been the Secretary-General of the European Society of Human Genetics (1989–1991) and the President of the German Society of Human Genetics (1990–1996), of which he became an honorary member in March 2011. He is a member of the American Society of Human Genetics, a corresponding member of the American College of Medical Genetics, a founding member of the European Society of Human Genetics and the Teratology Society, and several other scientific societies.

Table of Contents

Introduction 1

Chronology 19
Important Advances that Contributed to the Development of Genetics 19

Fundamentals 25

Prologue 26
Phylogenetic Tree of Living Organisms 26
Origins of Humans 28
Out of Africa: Toward Modern Humans 30
The Cell and Its Components 32
Genetic Background of Aging Processes 34
Molecular Basis of Genetics 36
Carbohydrates 36
Lipids (Fatty Acids) 38
Amino Acids 40
Nucleotides and Nucleic Acids 42
DNA and Its Components 44
DNA as a Carrier of Genetic Information 46
DNA Structure 48
DNA Replication 50
The Flow of Genetic Information: Transcription and Translation 52
Genetic Code 54
Eukaryotic Gene Structure 56
Analysis of DNA 58
Restriction Enzymes 58
DNA Amplification (PCR) 60
DNA Sequencing 62
Parallel DNA Sequencing (Next-Generation Sequencing) 64
DNA Cloning 66
DNA Libraries 68
Southern Blot Hybridization 70
Variability of DNA 72
DNA Variants 72
Genes and Mutation 74
Mutations Due to Base Modifications 76
Errors in Replication 78
Processing of DNA 80
DNA Repair Systems 80
Transposition 82
Trinucleotide Repeat Expansion 84
Eukaryotic Cells 86
Cell Communication 86
Haploid and Diploid Yeast Cells 88
Cell Cycle Control 90
Cell Division: Mitosis 92
Meiosis in Germ Cells 94
Meiosis Prophase I 96
Formation of Gametes 98
Programmed Cell Death 100
Cultured Cells 102
Formal Genetics 104
The Mendelian Traits 104
Transmission to the Next Generation 106
Independent Distribution 108
Phenotype and Genotype: Application in Genetic Counseling 110
Segregation of Parental Genotypes . . . 112
Monogenic Inheritance 114
Genetic Linkage and Recombination . 116
Genetic Linkage and Association 118
Quantitative Genetic Traits 120
Distribution of Alleles in a Population . . . 122
Hardy–Weinberg Equilibrium Principle . . . 124
Geographical Differences in Allelic Distribution 126
Inbreeding 128
Twins and Twinning 130
Chromosomes 132
Chromosomes and Genes 132
Chromosome Organization 134
Functional Elements of Chromosomes 136
Nucleosomes 138
Packing DNA in Chromosomes 140
The Telomere 142
Chromosomes in Metaphase 144
The Banding Patterns of Human Chromosomes 146

Karyotype of Man and Mouse 148
Preparation of Metaphase 150
Fluorescence *in Situ* Hybridization ... 152
Multicolor Fluorescence *In Situ* Hybridization (FISH) 154
Aneuploidy 156
Chromosome Translocation 158
Structural Chromosomal Aberrations . 160
Regulation of Gene Function 162
Ribosomal RNA and Protein Assembly 162
Stages of Transcription 164
Basic Principles of Gene Control 166
Regulation of Gene Expression in Eukaryotes 168
DNA–Protein Interactions 170
Other Forms of Transcription Control 172
Noncoding RNAs 174
Targeted Gene Disruption 176
Epigenetic Modifications 178
DNA Methylation 178
Reversible Changes in Chromatin 180
Genomic Imprinting 182
Mammalian X Chromosome Inactivation 184
Genetic Signal Pathways 186
Cellular Signal Transduction 186
Heterotrimeric G Proteins 188
TGF β and Wnt/β-Catenin Signaling Pathways 190
Hedgehog and TNF Signal Pathways .. 192
The Notch/Delta Signaling Pathway .. 194
Genes in Embryonic Development 196
Embryonic Development Genes in *Drosophila melanogaster* 196
Hox Genes 198
Zebrafish: A Translucent Vertebrate .. 200
Cell Lineage in a Nematode, *Caenorhabditis elegans* 202

Genomics 205

Genomics 206
Genomics: The Study of the Organization of Genomes 206
Genomes of Microorganisms 208
Architecture of the Human Genome .. 210
Regulatory Architecture of the Human Genome 212
Genome Analysis with DNA Microarrays 214
Genome Scan and Array–Comparative Genomic Hybridization 216
Comparative Genomic Hybridization 218
Genome-Wide Association Study 220
Mobile Genetic Elements 222
Genome Editing by the CRISPR-Cas System 224
Evolution of Genes and Genomes 226
Comparative Genomics 228
Genomic Structure of the Human X and Y Chromosomes 230
The Mitochondrial Genome of Man .. 232

Genetics in Medicine 235

Genetic Classification of Diseases 236
Genomic Disorders 236
Disorders due to Dysregulated Chromatin Structure 238
Disorders Resulting from Rearrangement of *Cis*-Regulatory Elements 240
Defects in Telomeres 242
Defective Lamins 244
Dysfunctional Cohesin 246
Dysfunctional Cilia (Ciliopathies) 248
Neurocristopathies 250
Dysregulated RAS-MAPK Signaling Pathway 252
Unstable Repeat Expansion 254
Fragile X Syndrome 256
Imprinting Disorders 258
Imbalanced Homeostasis 260
Mitochondrial Diseases 260
Chloride Channel Defects: Cystic Fibrosis 262
Genetic Defects in Ion Channels: LQT Syndromes 264
α_1-Antitrypsin Deficiency 266
Hemophilia A and B 268
von Willebrand Bleeding Disease 270
Pharmacogenetics 272
Cytochrome P450 (CYP) Genes 274
Metabolic Disorders 276
Genetics of Diabetes Mellitus 276
Amino Acid Degradation and Urea Cycle Disorders 278
Cholesterol Biosynthesis Pathway 280

Distal Cholesterol Biosynthesis Pathway 282
Familial Hypercholesterolemia 284
LDL Receptor Mutations 286
Lysosomal Storage Disorders 288
Lysosomal Enzyme Defects 290
Mucopolysaccharide Storage Diseases 292
Peroxisomal Disorders 294
Immune System 296
Components of the Immune System . 296
Immunoglobulin Molecules 298
Generation of Antibody Diversity 300
Immunoglobulin Gene Rearrangement 302
T-cell Receptor 304
The MHC Region 306
Evolution of the Immunoglobulin Superfamily 308
Primary Immunodeficiency Diseases . 310
Origins of Cancer 312
Genetic Causes of Cancer 312
Categories of Cancer Genes 314
Cancer Genomes 316
The *TP53* Tumor Suppressor Gene ... 318
The *APC* Gene and Polyposis Coli 320
Breast and Ovarian Cancer 322
Oncogenic Chromosome Translocations 324
Retinoblastoma 326
Neurofibromatosis 328
Genomic Instability Diseases 330
DNA Excision Repair Disorders 332
Impaired Cell and Tissue Structure 334
Cytoskeletal Proteins in Erythrocytes . 334
Hereditary Muscular Dystrophies 336
Duchenne's Muscular Dystrophy 338
FGF Receptor Mutations in Skeletal Dysplasias 340
Marfan's and Loeys–Dietz Syndromes . 342
Collagen Molecule Disorders 344
Osteogenesis Imperfecta 346
Molecular Basis of Bone Development 348
Hemoglobin Disorders 350
Hemoglobin 350
Hemoglobin Genes 352
Sickle Cell Disease 354
Mutations in Globin Genes 356
The Thalassemias 358
Hereditary Persistence of Fetal Hemoglobin (HPFH) 360
Sex Determination and Differentiation .. 362
Mammalian Sex Determination 362
Sex Differentiation 364
Disorders of Sexual Development 366
Congenital Adrenal Hyperplasia 368
Sensory Perception 370
Rhodopsin, a Photoreceptor 370
Pigmentary Retinal Degeneration 372
Color Vision 374
Auditory System 376
Odorant Receptors 378
Mammalian Taste Receptors 380
Chromosomal Aberrations 382
Numerical Chromosomal Aberrations 382
Triploidy, Monosomy X, Additional X or Y Chromosome 384
Microdeletion Syndromes 386
A Brief Guide to Genetic Diagnosis 388
A Brief Guide to Genetic Diagnosis ... 388
Gene and Stem Cell Therapy 390
Morbid Anatomy of the Human Genome 392
Chromosomal Locations of Human Genetic Diseases 392
Chromosomal Locations—Alphabetical List 398

Appendix 403

Glossary 421

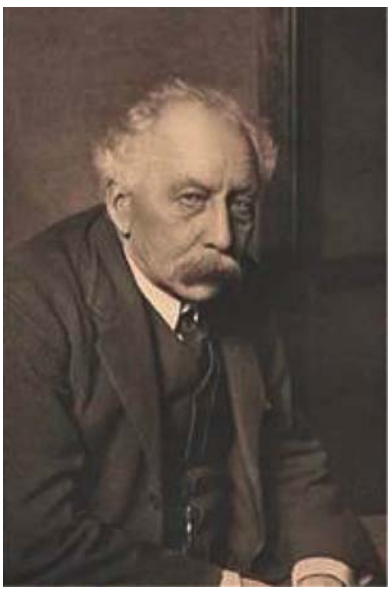
Index 447

Introduction

Genetics and Genomics

Scientific advances in genetics and genomics during the past 15 years have increased our knowledge about the structure and function of the processes carrying life at the molecular and cellular levels. These have yielded insights into a better understanding of the development and differentiation of cells and tissues, functional networks, distinguishing hereditary and nonhereditary influences on the causes of diseases and the differences between normal and cancer cells.

British biologist, William Bateson (1861–1926), proposed the term *genetics* for a new biological field devoted to the scientific investigation of heredity and variations in 1906. The new science of genetics provided a theoretical framework that could also be applied to the practice of plant and animal breeding. In the opening sentence of the *Mendel's Principles of Heredity*, published in 1909, Bateson stated, "Among the biological sciences the study of genetics occupies a central position." This was subsequently borne out by the long series of discoveries and technical advances that ensued and culminated in the completion of the Human Genome Project (HGP) in 2003 and the publication of a reference human genome sequence (IHGSC, 2001 and 2004). These unprecedented milestones provided information about the underlying biological structures and functions in genetics not available previously (see Part I, Fundamentals).



William Bateson
(1861–1926)

The term *genome* was introduced in 1920 by Hans Winkler (1877–1945) in Hamburg, from which the term *genomics* was derived. In addition, it was used for naming a new biomedical journal, *Genomics*, founded in 1987 by V. A. McKusick (1921–2010) and F. H. Ruddle (1929–2013). Genomics defines a novel field devoted to analyzing entire genomes rather than selected genes reflecting dramatic progress during the past two decades. The genome of animals, plants, and microorganism comprises the entirety of information required for life (see Part II, Genomics). Genomics integrates genetics, molecular biology, and cell biology and is concerned with all genes and their structure and function. According to the different scientific goals and methods employed in genomics, derivative terms are used, e.g., *transcriptome* for the analysis of all molecules involved in transcription and translation and their regulation; *proteome* for the analysis of all proteins that a cell or an organism produces; and *epigenome* for the analysis of nonhereditary (epigenetic) processes. Other areas are *functional genomics* (functional analysis), *comparative genomics* (establishing genomic maps with regard to the evolution of genomes), and *bioinformatics* (assembly, storage, and management of data).

Currently, genetics and genomics constitute a scientific area relevant to all fields of medical and biological disciplines, including anthropology, evolution, biochemistry, physiology, psychology, ecology, and their related fields of science. As both are theoretical and experimental science, genetics and genomics provide an understanding of the biology of the living world and genetic diseases (see Part III, Genetics in Medicine).

Genetic Basis of Life

Each of the approximately 80 trillion (10^{12}) cells of an adult human contains a program with life-sustaining information in its nucleus (except for red blood cells, which are devoid of nucleus). This information is contained in a longitudinal molecule, DNA (deoxyribonucleic acid, see later). The instructions are encoded in discrete units, the *genes*. Around 200 different types of cells perform the complex molecular transactions under the control of various genes. Genetic information enables the cells to convert atmospheric oxygen and ingested food into energy, regulate the synthesis and transport of biologically important molecules, protect themselves from harmful invaders, such as bacteria, fungi, and viruses by means of immune defense systems, and maintain the shape and mobility of bones, muscles, and skin. Genetically determined functions of the sensory organs enable us to see, hear, taste, and perceive heat, cold, and pain. Genetic information supports brain function with the ability to learn from experience, develop speech, and integrate the environmental input into cognate behavior. Reproduction and the detoxification of exogenous molecules are also under genetic control. In addition, structural modifications in the cell nucleus, not involving DNA directly contribute important regulatory functions (see epgenetics).

The living world consists of two basic types of cells, the smallest membrane-bound units capable of independent reproduction known as *prokaryotic* cells without a nucleus (represented by bacteria and archaea) and *eukaryotic* cells with a nucleus and complex internal structures in multicellular higher organisms. Genetic information is transferred from one cell to both the daughter cells formed at each cell division and from one generation to the next through specialized cells, the *germ cells, oocytes,* and *spermatozoa*.

Biological processes are mediated by biochemical reactions involving biomolecules called *proteins*. Genes contain encoded information for the intracellular synthesis of proteins. Each protein is made up of dozens to several hundreds of amino acids arranged in a linear sequence. This primary sequence of amino acids is called a *polypeptide,* which is subsequently folded into a specific three-dimensional structure, often in combination with other polypeptides, allowing biological function. All proteins that can be synthesized constitute the *proteome*. However, most cells do not produce all possible proteins, but rather a selection, depending on the type of cell. In addition to protein-coding genes, other genes encode information for the synthesis of small molecules made up of ribonucleic acid (RNA) required for regulatory functions.

Genetic information is stored in a linear fashion like a text of individual letters and words in a defined sequence that alone makes biological sense. This text consists of the nucleotide bases of DNA that is a read-only memory device of a genetic information system called the *genetic code*. In contrast to the binary system of strings of ones and zeros used in computers ("bits," which are then combined into eight binary digits, "bytes"), the genetic code in the living world uses a quaternary system of four nucleotide bases whose chemical names have the initial letters A, C, G, and T (see Part I, Fundamentals). The quaternary code used in living cells uses three building blocks, called a *triplet codon.* Each amino acid of a protein is encoded by a specific triplet. This genetic code is universal and is used by all living cells, including plants and viruses.

Genes

A gene is a unit of genetic information contained in a DNA segment. It is equivalent to a single sentence in a meaningful text. Thus, genetic information is highly analogous to a linear text and is amenable to being stored in computers. Depending on the organizational complexity of an organism, the number and size of genes vary considerably. Their number ranges from about 500 to 5,000 in most prokaryotes and about 6,000 to 40,000 in most eukaryotes. The minimal number of genes required to sustain independent cellular life is surprisingly small: around 250 for certain bacteria. Since many proteins are involved in related functions of the same pathway, they and their corresponding genes can be grouped into families of related function. It is estimated that approximately 21,000 protein-coding genes can be assigned to around 1,000 gene families (Demuth et al, 2006).

Genes are located on chromosomes in the nucleus of each cell. Chromosomes are individual, complex structures consisting of DNA and special

DNA-related proteins (histone proteins or simply histones). Chromosomes in eukaryotes come in homologous pairs, one of each pair derived from the mother and the other from the father. Humans have 23 pairs of chromosomes, consisting of chromosomes 1–22 and an X and a Y chromosome in males or two X chromosomes in females. The number and size of chromosomes vary in different organisms, but the total number of chromosomes is characteristic for each species. Genes are arranged linearly along each chromosome with each gene having a defined position, called a *gene locus*. In higher organisms, genes are structured into contiguous sections of coding and noncoding DNA sequences, called *exons* (coding) and *introns* (noncoding), respectively. Genes in multicellular organisms vary with respect to size (ranging from a few thousand to over a million nucleotide base pairs) and the number and size of exons. Each gene has regulatory DNA sequences, some of which act from a distance. The latter determine the state of activity of a gene called *gene expression*. Most genes in differentiated, specialized cells are permanently turned off. More than 90% of the 3 billion (3×10^9) base pairs of DNA in higher organisms do not carry known coding information but contain information with regulatory functions (see Part II, Genomics).

The linear text of information contained in the coding sequences of DNA in a gene cannot be read directly. Rather, its total sequence is first converted into a structurally related molecule with a corresponding sequence of codons. This process is called *transcription*. The entire set of molecules involved is called the *transcriptome*. The molecule resulting from transcription is known as ribonucleic acid (RNA) that serves as a template to arrange the amino acids into a polypeptide in the sequence specified by the genetic code. This process is called *translation*. Each of the 21 amino acids used by living organisms is encoded by a specific sequence of three RNA molecules.

Early Genetics between 1900 and 1910

Genetic information is transmitted from one generation to another according to defined rules known as Mendelian rules, named after the Augustinian monk, Gregor Mendel (1822–1884). When Mendel conducted crossbreeding experiments with garden peas in his monastery garden in Brünn (Brno, Czech Republic), he recognized that heredity was based on individual factors that were independent of each other (Mendel, 1866). These factors are transmitted from one plant generation to the next in a predictable pattern, each factor responsible for an observable trait. The trait one can observe is the *phenotype*. The underlying genetic information is the *genotype*.

Gregor Mendel
(1822–1884)

Gregor Mendel is credited with the discovery of basic principles of genetics and is regarded as the father of genetics. However, the fundamental importance of Mendel's conclusion was only recognized in 1900 independently by Correns, Tschermak, and De Vries. The term, *gene* for the type of heritable factors observed by Mendel was introduced in 1909 by the Danish biologist, Wilhelm Johannsen (1857–1927). Beginning in 1901, Mendelian inheritance was systematically analyzed in animals, plants, and humans. Some human diseases were recognized as having a hereditary cause. A

Thomas H. Morgan
(1866–1945)

form of brachydactyly (type A1, McKusick number OMIM 112500) observed in a large Pennsylvania sibship by W. C. Farabee (PhD thesis, Harvard University, 1903) was the first condition in humans to be described as being transmitted by autosomal dominant inheritance (Haws and McKusick, 1963).

Chromosomes were first observed in dividing cells in mitosis by Flemming in 1879 and in meiosis by Strasburger in 1888. Waldeyer coined the term, *chromosome* in 1888.

Theodor Boveri (1862–1915) recognized the genetic individuality of chromosomes in 1902. He wrote that not a particular number but a certain combination of chromosomes was necessary for normal development indicating that each individual chromosome possessed different genetic qualities.

Genetics became an independent scientific field in 1910 when Thomas H. Morgan introduced the fruit fly (*Drosophila melanogaster*) for systematic genetic studies at the Columbia University in New York. Among several other fundamental discoveries, Morgan and his coworkers demonstrated that genes were arranged on chromosomes in a sequential order. Morgan summarized this in 1915 as the *chromosome theory of inheritance* (Morgan, 1926). Although genetics was well established as a biological field by the end of the second decade of the last century, knowledge of the physical and chemical nature of genes was sorely lacking. However, their structure and function remained unknown.

Changes in Genes: Mutations

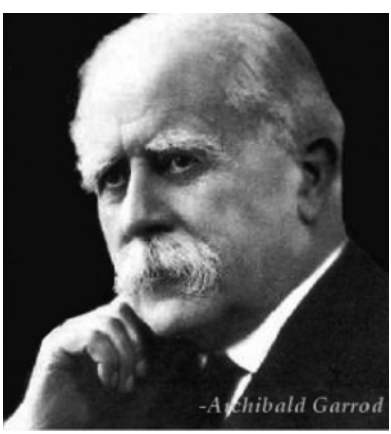
The integrity of the genetic program must be maintained without compromise; yet, it must be adaptable to respond to long-term changes in the environment during evolution. In 1901, H. de Vries recognized that genetic information was susceptible to changes. He introduced the term, *mutation* to describe the new observation. The systematic analysis of mutations contributed greatly to the further development of genetics. In 1927, H. J. Muller determined the spontaneous mutation rate in *Drosophila* and demonstrated that mutations could be induced by Rontgen rays. C. Auerbach and J. M. Robson in 1941 and F. Oehlkers independently in 1943 observed that certain chemical substances could also induce mutations. However, it remained unclear what a mutation actually was, since the physical basis for the transfer of genetic information was not known. Errors in maintaining and transmitting genetic information occur in all living systems.

Genes of fundamental importance do not tolerate changes (mutations) that compromise function. As a result, deleterious mutations do not accumulate in any substantial number. Elaborate cellular systems can recognize and eliminate faults in the integrity of DNA and genes (*DNA repair*).

Genetic Individuality

The sequence of DNA differs among unrelated individuals. This is referred to as *DNA polymorphism*. Usually, this involves just one nucleotide (*single nucleotide polymorphism*, SNP). Other forms of DNA polymorphism involve small or large blocks of repeated nucleotide sequences (*copy number variation*, CNV). Such individual genetic differences form the basis of genetic individuality. In 1908, the English mathematician, G. H. Hardy and the German physician, W. Weinberg independently recognized that Mendelian inheritance accounted for regularities in the distribution of genetic variants in different populations.

In 1902, Archibald Garrod (1857–1936), who served as Regius Professor of Medicine at the Oxford University, demonstrated that four congenital metabolic diseases (albinism, alkaptonuria, cystinuria, and pentosuria) were transmitted by autosomal recessive inheritance. He called these *inborn errors of metabolism*. In addition, Garrod was the first to recognize that subtle biochemical differences among individuals resulted from individual genetic differences. In 1931, he published a prescient monograph entitled *The Inborn Factors in Disease* (Garrod, 1931). He suggested that small genetic differences might contribute to the causes of diseases. Garrod, together with W. Bateson, introduced genetic concepts into medicine in the early years of genetics between 1902 and 1909. In late 1901, Garrod and Bateson began an extensive correspondence about the genetics of alkaptonuria and the significance of consanguinity, which Garrod had observed among the parents of affected individuals. In a letter to Bateson on January 11, 1902, Garrod wrote, "I have for some time been collecting information as to specific and individual differences of metabolism, which seems to me a little explored but promising field in relation to natural selection, and I believe that no two individuals are exactly alike chemically any more than structurally"



Archibald Garrod
(1857–1936)

(Bearn, 1993). However, Garrod's concept of genetic individuality of man was not recognized at the time. One reason might have been that the lack of knowledge on the structure and function of genes despite early fundamental discoveries. At present, we recognize individual susceptibility to a disease as an important factor in its causes (see Part III, Genetics in Medicine). Today, genetic individuality has been defined by structural DNA variants in 26 human populations (Sudmant et al, 2015).

Rise of Modern Genetics between 1940 and 1953

With the demonstration in the fungus, *Neurospora crassa* that one gene was responsible for the formation of one enzyme ("one gene, one enzyme" as formulated by Beadle and Tatum in 1941), the relationship between genetics and biochemistry became apparent. Systematic studies in microorganisms led to other important advances in the 1940s. Bacterial genetics began in 1943 when Salvador E. Luria and Max Delbrück discovered mutations in bacteria. Other important advances were the demonstration of genetic recombination demonstrated in bacteria by Lederberg and Tatum in 1946 and in viruses by Delbrück and Bailey in 1947 as well as the observation of spontaneous mutations in bacteriophages by Hershey in 1947. The study of genetic phenomena in microorganisms turned out to be as significant for the further development of genetics as the analysis of *Drosophila* had been 35 years earlier (Cairns et al, 1978). A very influential, small book entitled *What is Life?* by the physicist, E. Schrödinger (1944) postulated a molecular basis for genes. Henceforth, the elucidation of the molecular biology of the gene became a central theme in genetics.

Genetics and DNA

A major discovery by Avery, MacLeod, and McCarty, at the Rockefeller Institute in New York in 1944, indicated that DNA carried genetic information in bacteria. DNA was recognized as a chemically relatively simple, long-chained molecule by Friedrich Miescher in 1869 but considered too simple for genetic information. In 1928, F. Griffith observed that permanent (genetic) changes could be induced in pneumococcal bacteria by a cell-free extract derived from other strains of pneumococci,

called the *transforming principle*. Avery and his coworkers demonstrated DNA to be the transforming principle. In 1952, Hershey and Chase proved that DNA alone carried genetic information and excluded other molecules. With this discovery, the question of the structure of DNA took center stage in biology as described by McCarty (1985) and Dubos (1976).

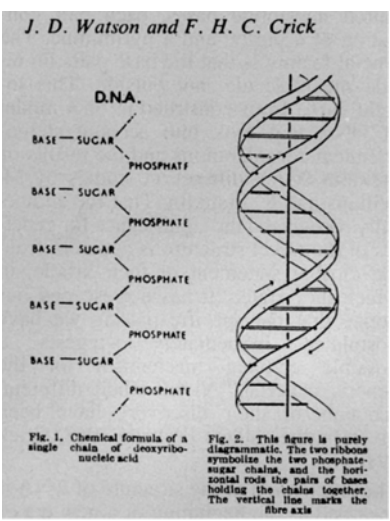
This question regarding the structure of DNA was resolved in a short, one-page article in the journal, *Nature* on April 25, 1953 (Watson and Crick, 1953). The authors proposed the structure of DNA as a double helix, which consisted of two complementary chains of alternating sugar (deoxyribose) and monophosphate molecules, oriented in opposite directions. Inside the helical molecule are the paired nucleotide bases. Each pair consists of a pyrimidine and a purine, either of a cytosine (C) and a guanine (G) or of a thymine (T) and adenine (A). The crucial feature is that the base pairs (C-G and A-T) are located inside the molecule, not outside. The DNA structure as a double helix was derived from model building using the. This idea was largely supported by an X-ray diffraction photograph of crystalline DNA obtained by Rosalind E. Franklin (Franklin and Gosling, 1953), indicating DNA to be a helix (Maddox, 2002).

The structure of DNA as a double helix with the nucleotide bases inside explains two fundamental genetic mechanisms: the storage of genetic information in a linear, readable pattern and the replication of genetic information to ensure its accurate transmission from one generation to another.

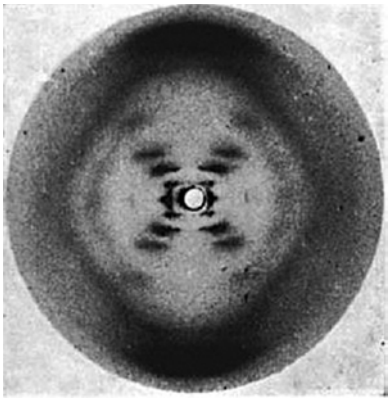
Two publications accompanied the article by Watson and Crick (1953) describing additional aspects of the DNA structure (Wilkins et al, 1953; Franklin and Gosling, 1953). An earlier basis for recognizing the structure of DNA was the discovery by E. Chargaff in 1950 who demonstrated that cytosine and guanine as well as adenine and thymine were present in the same quantity in DNA. However, this was not recognized as a result of pairing (Wilkins, 2003). Vivid, albeit different, accounts of the discovery of the structure of DNA have been provided by the scientists involved (Watson, 1968; Crick, 1988; Wilkins, 2003; Maddox, 2002 about Rosalind Franklin).

The elucidation of the structure of DNA is regarded as the beginning of a new era of

Oswald T. Avery in 1937
(1877–1955)



DNA structure 1953



Franklin's photograph 51 indicating DNA to be a helix.

molecular biology and genetics. The description of DNA as a double-helix structure led directly to an understanding of the possible structure of genetic information. When F. Sanger determined the sequence of amino acids of insulin in 1955, he provided the first proof of the primary structure of a protein. This showed that the sequence of amino acids in proteins corresponded to the sequential character of DNA. The genetic code required for the synthesis of proteins from DNA and mRNA was determined in the years from

Rosalind Franklin (1920–1958)

Watson and Crick in 1953
(Photograph by Anthony Barrington Brown, Nature 421: 417, 2003)

Maurice Wilkins (1916–2004)

1963 to 1966 by Nirenberg, Mathaei, Ochoa, Benzer, Khorana, and others. Several authors have presented detailed accounts of these developments (Watson, 1968, 2000; Chargaff, 1978; Stent, 1981; Watson and Tooze, 1981; Crick, 1988; Judson, 1996; Wilkins, 2003).
With the structure of DNA known, the nature of the gene could be redefined in molecular terms. In 1955, Seymour Benzer (1921–2007) established the first genetic fine structure. He

established a genetic map of contiguous deletions of a region (rII) of the bacteriophage T4. He found that mutations could be divided into two functional groups: A and B. Mutants belonging to different groups could complement each other by eliminating the effects of the deletion; those belonging to the same group could not. This work defined the gene in terms of molecular function.

New Methods in the Development of Genetics after 1953

From the beginning, genetics has been a field in which new concepts are based on the development of new experimental methods. In the 1950s and 1960s, the groundwork was laid for *biochemical genetics* and *immunogenetics*. Relatively simple but reliable procedures for separating complex molecules by different forms of electrophoresis, methods of synthesizing DNA in vitro by Kornberg in 1956, and other approaches were applied to genetics. The introduction of cell culture methods was of particular importance for the genetic analysis of humans. G. Pontecorvo introduced the genetic analysis of cultured eukaryotic cells (*somatic cell genetics*) in 1958. The study of mammalian genetics, with increasing significance for studying human genes, was facilitated by methods for fusing cells in culture (*cell hybridization*) introduced by T. Puck, G. Barski, and B. Ephrussi in 1961 and the development of a cell culture medium for selecting certain mutants in cultured cells (hypoxanthine–aminopterin–thymidine [HAT] *medium*) by J. Littlefield in 1964. The genetic approach that had been successful in bacteria and viruses could now be applied to higher organisms, thus avoiding the obstacles of a long generation time and breeding experiments. A hereditary metabolic defect in humans (galactosemia) was demonstrated for the first time in cultured human cells in 1961 by R. S. Krooth. The correct number of chromosomes in humans was determined independently in 1956 by Tjio and Levan and by Ford and Hamerton. Lymphocyte cultures were introduced for chromosomal analysis by Hungerford, and Nowell and coworkers in 1960. The first chromosomal aberrations in humans were described in 1959. The replication pattern of human chromosomes was described by German in 1962. These and other developments paved the way for a new field, *human genetics*.

Molecular Genetics

From around 1970 onward, genetics developed a new molecular dimension based on new techniques, which allowed the analysis of DNA directly. It became possible to determine the sequence of the DNA nucleotide bases by methods developed in 1977 by F. Sanger and Maxam and Gilbert (DNA sequencing). Even small amounts of DNA could be multiplied by a polymerase chain reaction (PCR) introduced in 1985. Today, molecular DNA analysis has been replaced by automated procedures allowing a high throughput analysis in a few days of what used to take weeks and longer, and at lower costs.

The discovery of reverse transcriptase, independently by H. Temin and D. Baltimore in 1970, upset the central dogma in genetics that the flow of genetic information was in one direction only, from DNA to RNA and from RNA to a protein as the gene product. *Reverse transcriptase* is an enzyme complex in RNA viruses (*retroviruses*) that transcribes RNA into DNA. Apart from being an important biological finding, such an enzyme can be utilized to obtain *complementary DNA* (cDNA) that corresponds to the coding regions of an active gene. Thus, a gene can be analyzed directly without knowledge of its gene product. Enzymes that cleave DNA at specific sites, *restriction endonucleases* or simply *restriction enzymes*, were discovered in bacteria by W. Arber in 1969 and by D. Nathans and H. O. Smith in 1971 (restriction analysis). Using these enzymes, DNA fragments of reproducible and defined sizes can be obtained and selected regions of a DNA molecule analyzed. DNA fragments of different origins could be joined and their properties analyzed. All these methods are collectively referred to as *recombinant DNA technology* (see Part I, Fundamentals).

In 1977, recombinant DNA analysis led to the unexpected finding that genes in higher organisms are not continuous segments of coding DNA but are interrupted by noncoding segments. The size and pattern of coding DNA segments, called *exons* and of the noncoding segments, called *introns* (two new terms

introduced by W. Gilbert in 1978) are characteristic for each gene, known as the *exon/intron* structure.

Genes and Evolution

The evolutionary biologist, Theodosius Dobzhansky at the Rockefeller University had stated, "Nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973) at a time when a relationship between genetics and evolution was not yet generally accepted. Today, one could say, "Nothing in evolution makes sense except in the light of genetics." Genes with comparable functions in different organisms share structural features. Occasionally, these are nearly identical, which is attributed to the process of evolution. Living organisms are related to each other by their origin from a common ancestor. Genes evolve within the context of the genome of which they constitute a part. An important evolutionary mechanism is the duplication of a gene or other DNA sequences (Ohno, 1970). During the course of evolution, existing genes or parts of genes are duplicated and reshuffled and brought together in new combinations. The human genome contains multiple sites that were duplicated during evolution (see Part II, Genomics). Most genes arise during evolution from preexisting ones or parts of genes existing before.

Transposable DNA

Certain DNA sequences can change their location by moving to a new site. Several mechanisms exist, collectively called *transposition*. This was first described between 1950 and 1953 by Barbara McClintock at the Cold Spring Harbor Laboratory, New York. She described genetic changes in Indian corn plants (maize) and their effect on the phenotype induced by a mutation in a gene that was not located at the site of the mutation. Surprisingly, such a gene could exert a type of remote control. In subsequent work, McClintock described the special properties of this group of genes, which she called *controlling genetic elements*. Different controlling elements could be distinguished according to their effects on other genes and the mutations caused. Originally, her work was received with skepticism (Fox Keller, 1983; Fedoroff and Botstein, 1992). In 1983, she received the Nobel Prize (McClintock, 1984). Today, we know that different types of transposons with different mechanisms form families of transposons. Transposition lends the genome flexibility during the course of evolution. Occasionally, a transposon inserts itself into a gene and causes a disease (Reilly et al, 2913; see Part II, Genomics).

Epigenetics

The term, epigenetics refers to a branch of biology aimed at studying the causal interactions between genes and gene products (proteins and small RNA molecules involved in the regulatory processes), which result in a phenotype. Epigenetics has attracted considerable interest in recent years. In 1942, C. H. Waddington derived the term from the words, genetics and epigenesis. It brings genetics and developmental biology together by focusing on heritable changes in gene expression without concomitant changes in the DNA sequence. Epigenetic changes are important mechanisms for control of genetic activity of many genes or groups of genes.

DNA-associated proteins (histone proteins or for short, histones) in the chromatin (the packaged DNA in the cell nucleus) are modified by different molecular mechanisms. Special enzymes add or remove methyl groups, acetyl groups, or phosphate groups at specific sites. This alters the functional state in chromatin (see p. 180 and 238). Certain states are associated with genetic activity, whereas others represent a repressed genetic state (inactivity). More than 250 differentially methylated regions (DMRs) in the genomes of human and mouse depict a specific pattern of DNA methylation.

Methylated DNA is associated with a genetically inactive state, whereas unmethylated DNA is found in genetically active regions. With certain genes, only one allele is expressed, either the one of maternal (mat) or the one of paternal (pat) origin.

Here, only one allele of a given gene or region is unmethylated and active, whereas the other allele is methylated and inactive. The methylation pattern is determined by the parental origin of the allele. Thus, either the allele of paternal origin (pat) or the allele of maternal origin (mat) is methylated. This pattern, called *genomic imprinting*, is transmitted to daughter cells and maintained. DNA methylation is an important control mechanism in gene expression such that errors in establishing or maintaining the correct methylation pattern result

in imprinting disorders (pp.194 and 368 in CAG4e).

Genetic Classification of Diseases

Modern genetic and genomic analysis immensely contributes to the diagnosis and management of human diseases (human genetics). Arguably, human genetics was inaugurated when The American Society of Human Genetics and the first journal of human genetics, *The American Journal of Human Genetics* were established in 1949. In addition, the first textbook of human genetics appeared in 1949, Curt Stern's *Principles of Human Genetics* (Stern, 1973).

As outlined in detail by Barton Childs (1999 and 2016; Childs and Pyeritz, 2013), two different views of the concept of disease can be distinguished. One, first introduced by William Osler in his fundamental textbook, *Principles and Practice in Medicine* in 1892, views a disease as a "broken machine" that needs to be recognized and repaired. In this system, diseases are mainly classified according to their phenotype, i.e., the manifestation according to organ systems, age, and gender. It does not ask why a particular disease affects one individual and not another. In contrast, Garrod's concept of genetic individuality poses the question of why a particular disease occurs. The Garrodian view of disease considers disease as a consequence of an imbalance within a patient's genetic individuality and with environmental living conditions. In human genetics, diseases are classified according to the genotype rather than the phenotype (clinical manifestation). Here, causal changes at gene loci and in genes primarily define a disease rather than the phenotype. The types of mutations represent a *molecular pathology*. Many genetic diseases have a similar phenotype, although they result from pathological changes in different genes. This is referred to as *etiological* (genetic) *heterogeneity*. Furthermore, rearrangements at different sites in one and the same gene may result in different phenotypes. Genetic heterogeneity is an important principle that always needs to be considered in the diagnosis of human genetic disorders.

A disease is genetically determined if it is mainly or exclusively caused by a functional failure in genes or their regulation. Genetic disorders can be assigned to six broad categories: (1) monogenic, (2) chromosomal, (3) complex (multigenic with interaction with environmental influences), (4) genomic disorders resulting from certain structural features of the human genome that predispose to disease-causing rearrangements of DNA segments, (5) somatic mutations (different forms of cancer), and (6) imprinting disorders resulting from aberrant patterns of imprinted genes (see epigenetics).

Several disorders can be grouped according to a signal pathway that is interrupted by a mutation or a rearrangement (see Part III, Genetics in Medicine). The estimated overall frequency of genetically determined diseases of different categories in the general population is about 3 to 5% (see Table).

Table Categories and frequency of genetically determined diseases

Category of disease	Estimated frequency per 1,000 individuals[a]
Monogenic diseases total	5–17
Autosomal recessive	2–7
Autosomal dominant	2–8
X-chromosomal	1–2
Chromosome aberrations (light microscopy)	5–7
Complex disorders (multigenic)	70–90
Genomic disorders	5–10
Somatic mutations (cancer)	200–250
Mitochondrial disorders	2–5
Imprinting disorders	1–2

[a] Approximate estimates based on various sources.

The most important and frequent group of diseases is the group comprising multigenic or multifactorial diseases. These result from environmental influences interacting with the individual genetic makeup of the affected individual. Important examples are relatively common chronic diseases, such as high blood pressure, hyperlipidemia, diabetes mellitus, gout, psychiatric disorders, disorders with intellectual impairment, aging disorders, and certain congenital malformations. Their cause is not a mutation in a single gene, but rather specific variants in several genes with predisposition to a particular disorder. Another common category is cancer, a large, heterogeneous group of nonhereditary genetic disorders resulting from mutations in somatic cells or hereditary changes in germ cells.

Numerous subspecialties of human genetics have arisen, such as *biochemical genetics*, *immunogenetics*, *somatic cell genetics*, *cytogenetics*, *clinical genetics*, *population genetics*, *teratology*, mutational studies, and others. The development of human genetics has been well summarized by McKusick (1992), McKusick and Harper (2013), and Vogel and Motulsky (1997).

The enormous progress in the medical aspects of human genetics (*medical genetics*), in particular for monogenic disorders, is best documented in *Mendelian Inheritance in Man* (MIM), a catalog of human genes and genetic disorders (McKusick, 1998). It is freely available online: *Online Mendelian Inheritance in Man* (OMIM). It was first established in 1966 by Victor A. McKusick (1921–2008) at the Johns Hopkins University in Baltimore and went through 12 printed editions (1968–1998, see p. 398 in CAG). Each entry carries a six-digit number. The first digit indicates the mode of inheritance or status of molecular knowledge (1, autosomal dominant; 2, autosomal recessive; 3, X-chromosomal; 4, Y-chromosomal; 5, mitochondrial; 6, additional molecular information [OMIM, Online Mendelian Inheritance in Man, see p. 392]). All diseases mentioned are designated with their 6-digit OMIM number throughout this book.

The Human Genome Project and other International Initiatives

A new dimension in genomics was introduced into biomedical research in 1990 by the Human Genome Project (HGP) and related programs in many other organisms (Lander and Weinberg, 2000; Green and Guyer, 2011). It ended in 2003 with the publication of the sequence of human DNA in a reference sequence (IHGSC, 2004). The HGP was an international organization representing several countries under the leadership of biomedical centers in the United States and the United Kingdom. The main goal of the HGP was to determine the entire sequence of the 3 billion nucleotide pairs in the DNA of the human genome. At the time, this was a daunting task, as it was comparable to deciphering each individual 1-mm wide letter of a strip of text 3,000 km long. The first draft of a sequenced human genome covering approximately 90% of the genome was announced in June 2000 (IHGSC, 2001; Venter et al, 2001). The complete DNA sequence of human genome was published in 2004 (IHGSC, 2004). All human chromosomes have been sequenced (see Selected Websites for Access to Genetic and Genomic Information: Nature Web Focus: Human Genome Collection and OMIM). Several other international

Victor A. McKusick (1921–2010) (www.hopkinsmedicine.org)

initiatives have formed to investigate a defined area of research. Two examples are the HapMap Project and ENCODE.

The International HapMap Project was initiated in 2002 as an international multi country project aimed at identifying all individual genetic variants in the DNA sequence. These variants may have important influences on the causes of diseases or responses to therapeutic drugs (see Website for Access to Genetic and Genomic Information: International HapMap Project).

ENCODE (Encyclopedia of DNA Elements) is an international consortium studying all functional elements in the human genome (see Selected Websites for Access to Genetic and Genomic Information: Encyclopedia of DNA Elements).

The Human Epigenome Project (HEP) aims to identify and catalog all variable methylation positions in the human genome.

A Misconception in Genetics: Eugenics

Eugenics, a term coined by Francis Galton in 1882, aims to improve humans by genetic means. Between about 1900 and 1935, many countries adopted policies and laws that were assumed to reduce or eliminate an accumulation of "undesired" genetic traits in a population. It was believed that the "White race" was superior to others, but proponents did not realize that genetically defined human races do not exist. Eugenics assumed that sterilizing individuals with diseases considered hereditary would improve the human society. By 1935, sterilization laws had been passed in Denmark, Norway, Sweden, Germany, and Switzerland, as well as in 27 states of the United States. Individuals with mental impairment of variable degree or epilepsy and criminals and homosexuals were the prime targets. Although in most cases, the stated purpose was eugenic, sterilizations were performed for social rather than genetic reasons.

The complete lack of knowledge of the structure and function of genes might have contributed to the eugenic misconceptions, which assumed that "bad genes" could be eliminated from human populations. However, the disorders targeted are either not hereditary or have a complex genetic background. Sterilization simply will not reduce the frequency of genes contributing to mental retardation and other disorders. In Nazi Germany from 1933 until the end of World War II in 1945, eugenics was used as a pretext for widespread discrimination and the murder of millions of innocent human beings claimed to be "worthless" (Müller-Hill, 1988; Vogel and Motulsky, 1997; Strong, 2003). However, all such reasons supposedly based on genetics have no scientific basis. Modern genetics has shown that the ill-conceived eugenic approach cannot eliminate or reduce the frequency of human genetic diseases. Thus, incomplete genetic knowledge was applied to human individuals at a time when nothing was known about the structure of genes. Indeed, up to 1949, no fundamental advances in genetics had been obtained by studies in humans. Quite the opposite holds true today. It is evident that genetically determined diseases cannot be eradicated and the society has to adjust to their occurrence. No one is free from a genetic burden. Every individual carries approximately 10 or more potentially harmful changes in the genome that under certain circumstances could manifest as a genetic disease unexpectedly in any family.

Ethical and Societal Issues, Education

The HGP also devoted attention and resources to ethical, legal, and social issues (the Ethical, Legal and Social Implications [ELSI] Research program). This constituted a significant part of the HGP, in view of the far-reaching consequences of the current and expected knowledge about human genes and the genome. Depending on the family history and the type of disease, it is now often possible to obtain diagnostic information about a disease years or even decades prior to its manifestation. This widens the time frame of a diagnosis. Furthermore, not only the affected individual, the patient, but also other, i.e., the unaffected family members may seek information about their own risk for a disease or the risk for a disease in their offspring. The ability to perform presymptomatic or predictive genetic testing raises new questions about the use of genetic data. The decision to perform a genetic test has to take into account a person's view on an individual basis, and informed consent be obtained only after the individuals involved have been properly counseled about the purpose, validity, and reliability, and the possible consequences of the test result. In some countries, laws have been introduced to ensure that

any genetic information generated is used in the best interests of the individual involved, informed consent is obtained, and the confidentiality of data is assured.

The completion of the HGP and the introduction of new tools for genomic research, in particular the relatively inexpensive high-capacity methods of sequencing DNA (massive parallel sequencing, "second generation," p. 64), since 2005 have ushered in a new era of genomic medicine (Green and Guyer, 2011; Lupski et al, 2011). In addition, the whole genome can now be subjected to a search for genetic factors that contribute to the causes of a given disorder (genome-wide association studies, p. 220). Genetic and nongenetic bases of disease can be distinguished and individual risk factors determined. Individual adverse responses to therapeutic agents (pharmacogenetics) can now be defined at the level of the whole genome (pharmacogenomics). Any new, often unsuspected, genomic information has to be channeled into individual counseling and decision making.

Literature Cited in the Text

Bateson W. Mendel's Principles of Heredity. Cambridge: University of Cambridge Press; 1909

Bearn AG. Archibald Garrod and the Individuality of Man. Oxford: Oxford University Press, Oxford; 1993

Cairns J, Stent GS, Watson JD, eds. Phage and the Origins of Molecular Biology. New York: Cold Spring Harbor Laboratory Press; 1978

Chargaff E. Heraclitean Fire: Sketches from a Life before Nature. New York: Rockefeller University Press; 1978

Childs B. Genetic Medicine. A Logic of Disease. Baltimore: Johns Hopkins University Press; 1999

Childs B. A Logic of Disease. The Online Metabolic and Molecular Bases of Inherited Disease. 6th ed. Elsevier; Chap 2

Childs B, Pyeritz RE. Medicine in a genetic context. In: Rimoin DL, Pyeritz R, Korf B, eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York: Elsevier; 2013. Chap 2

Crick F. What Mad Pursuit: A Personal View of Scientific Discovery. New York: Basic Books; 1988

Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. The evolution of mammalian gene families. PLoS One 2006;1:e85 PubMed

Dobzhansky T. Nothing in biology makes sense except in the light of evolution. Am Biol Teach 1973;35(3):125–129

Dubos RJ. The Professor, the Institute, and DNA: Oswald T. Avery, his Life and Scientific Achievements. New York: Rockefeller University Press; 1976

Fedoroff N, Botstein D, eds. The Dynamic Genome: Barbara McClintock's Ideas in the Century of Genetics. New York: Cold Spring Harbor Laboratory Press; 1992

Fox Keller EA. A Feeling for the Organism: the Life and Work of Barbara McClintock. New York: W. H. Freeman; 1983

Franklin RE, Gosling RG. Molecular configuration in sodium thymonucleate. Nature 1953;171(4356): 740–741 PubMed

Garrod AE. The Inborn Factors in Disease: an Essay. Oxford: Clarendon Press; 1931

Green ED, Guyer MS; National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. Nature 2011;470(7333):204–213 PubMed

Haws DV, McKusick VA. Farabee's brachydactylous kindred revisited. Bull Johns Hopkins Hosp 1963;113:20–30 PubMed

Lander ES, Linton LM, Birren B, et al; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860–921 PubMed

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431(7011):931–945 PubMed

Judson HF. The Eighth Day of Creation. Makers of the Revolution in Biology, expanded edition. New York: Cold Spring Harbor Laboratory Press; 1996

Lander ES, Weinberg RA. Genomics: journey to the center of biology. Science 2000;287(5459):1777–1782 PubMed

Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell 2011;147(1):32–43 PubMed

Maddox B. Rosalind Franklin. Dark Lady of DNA. New York: HarperCollins; 2002

McCarty M. The Transforming Principle. New York: W. W. Norton; 1985

McClintock B. The significance of responses of the genome to challenge. Science 1984;226(4676):792–801 PubMed

McKusick VA. Human genetics: the last 35 years, the present, and the future. Am J Hum Genet 1992;50 (4):663–670 PubMed

McKusick VA, Harper P. History of medical genetics. In: Rimoin D, Pyeritz R, Korf B, eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York: Elsevier; 2013. Chap 1

McKusick VA. Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders. 12th ed. Baltimore: Johns Hopkins University Press; 1998

Mendel G. Versuche über Pflanzen-Hybriden. Verh Naturforsch Ver Brünn 1866;4: 3–47

Morgan TH. The Theory of the Gene. Enlarged and Revised Edition. New Haven: Yale University Press; 1926

Müller-Hill B. Murderous Science. Oxford: Oxford University Press; 1988

Ohno S. Evolution by Gene Duplication. New York, Berlin Heidelberg: Springer-Verlag; 1970

OMIM. Online Inheritance in Man. A Catalog of Human Genes and Genetic Disorders. Available at: http://www.ncbi.nlm.nih.gov/omim

Reilly MT, Faulkner GJ, Dubnau J, Ponomarev I, Gage FH. The role of transposable elements in health and diseases of the central nervous system. J Neurosci 2013;33(45):17577–17586 PubMed

Schrödinger E. What Is Life? The Physical Aspect of the Living Cell. New York: Penguin Books; 1944

Watson JD. The Double Helix: A Personal Account of the Discovery of the Structure of DNA. Stent GS, ed. London: Weidenfeld & Nicolson; 1981

Stern C. Principles of Human Genetics. 3rd ed. San Francisco: W. H. Freeman; 1973

Strong C. Eugenics. In: Cooper DV, ed. Encyclopedia of the Human Genome. Vol. 2. London: Nature Publishing Group; 2003:335–340

Sudmant PH, Rausch T, Gardner EJ, et al; 1000 Genomes Project Consortium. An integrated map of structural variation in 2,504 human genomes. Nature 2015;526(7571):75–81 PubMed

Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science 2001;291(5507):1304–1351 PubMed

Vogel F, Motulsky AG. Human Genetics: Problems and Approaches. 3rd ed. Heidelberg: Springer-Verlag; 1997

Watson JD. The Double Helix. A Personal Account of the Discovery of the Structure of DNA. New York: Atheneum; 1968

Watson JD. A Passion for DNA. Genes, Genomes, and Society. New York: Cold Spring Harbor Laboratory Press; 2000

Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953;171(4356):737–738 PubMed

Watson JD, Tooze J. The DNA Story: a documentary history of gene cloning. San Francisco: W. H. Freeman; 1981

Wilkins MH, Stokes AR, Wilson HR. Molecular structure of deoxypentose nucleic acids. Nature 1953;171(4356):738–740 PubMed

Wilkins M. The Third Man of the Double Helix. Oxford: Oxford University Press; 2003

(revised 8 June 2017)

Further Reading

Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 6th ed. New York: Garland Publishing Co; 2015

Brown TA. Genomes. 3rd ed. New York: Garland Science; 2007

Dobzhansky T. Genetics of the Evolutionary Process. New York: Columbia University Press; 1970

Dunn LC. A Short History of Genetics. New York: McGraw-Hill; 1965

Erickson RP, Wynshaw-Boris AJ, eds. Epstein's Inborn Errors of Development: The Molecular Basis of Clinical Disorders of Morphogenesis. 3rd ed. Oxford: Oxford University Press; 2016

Gilbert SF, Barresi MJF. Developmental Biology. 11th ed. Sunderland: Sinauer; 2016

Harper PS. Practical Genetic Counselling. 7th ed. London: Edward Arnold; 2010

Griffiths AJF, Wessler SR, Lewontin RC. Carroll SB. Introduction to Genetic Analysis. 10th ed. San Francisco: W. H. Freeman; 2014

Domchek SM, Jameson JL, Miesfeldt S. The Practice of Genetics in Clinical Medicine. Harrison's Principles of Internal Medicine. 10th ed. New York: McGraw-Hill Medical; 2015. Chap 84

Jobling MA, Hollox E, Hurles M, Kivisild T, Tyler-Smith C. Human Evolutionary Genetics. 2nd ed. New York: Garland Science; 2014

Jorde LB, Carey JC, Bamshad MJ. Medical Genetics. 5th ed. Philadelphia: Elsevier; 2015

King R, Rotter J, Motulsky AG, eds. The Genetic Basis of Common Disorders. 2nd ed. Oxford: Oxford University Press; 2002

King RC, Mulligan PK, Stansfield WD. A Dictionary of Genetics. 8th ed. Oxford: Oxford University Press; 2013

Jan K, Naoyuki T. Where Do We Come From? The Molecular Evidence for Human Descent. Heidelberg: Springer-Verlag; 2002

Krebs JE, Kilpatrick ST, Goldstin ES. Lewin's Genes XI. Sudbury: Jones & Bartlett; 2013

Lodish H, Berk A, Kaiser CA, et al. Molecular Cell Biology. 8th ed. New York: W. H. Freeman; 2016

Lupski JR, Stankiewicz P, eds. Genomic Disorders. The Genomic Basis of Disease. Totowa: Humana Press; 2006

McCLINTOCK B. Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol 1951;16:13–47 PubMed

McClintock B. Induction of instability at selected loci in maize. Genetics 1953;38(6):579–599 PubMed

Morgan TH. The Theory of the Gene. New Haven: Yale University Press; 1926

Nussbaum RL, McInnes RR, Willard HF. Thompson & Thompson Genetics in Medicine. 8th ed. Philadelphia: W. B. Saunders; 2016

Rimoin DL, Connor JM, Pyeritz RE, Korf BR, eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 7th ed. Edinburgh: Elsevier; 2013

Scriver CR, Beaudet AL, Sly W, Valle D, eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York: McGraw-Hill; 2001. Available at: http://www.ommbid.org. Accessed June 6, 2017

Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med 2010;61:437–455 PubMed

Speicher MR, Antnonarakis SE, Motulsky AG, eds. Vogel and Motulsky's Human Genetics. Problems and Approaches. 4th ed. Heidelberg: Springer-Verlag; 2010

Stebbins GL. Darwin to DNA. Molecules to Humanity. San Francisco: W. H. Freeman; 1982

Stent G, Calendar R. Molecular Genetics. An Introductory Narrative. 2nd ed. San Francisco: W. H. Freeman; 1978

Strachan T, Goodship J. Chinnery P Genetics and Genomics in Medicine. London: Garland Science; 2015

Sturtevant AH. A History of Genetics. New York: Harper & Row; 1965

Turnpenny PD, Ellard S. Emery's Elements of Medical Genetics. 14th ed. Edinburgh, Philadelphia: Elsevier-Churchill Livingstone; 2011

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. Molecular Biology of the Gene. 6th ed. New York: Pearson/Benjamin Cummings and Cold Spring Harbor Laboratory Press; 2008

Weinberg RA. The Biology of Cancer. 2nd ed. New York: Garland Science; 2013

Weatherall DJ. The New Genetics and Clinical Practice. 3rd ed. Oxford: Oxford University Press; 1991

Whitehouse HLK. Towards an Understanding of the Mechanism of Heredity. 3rd ed. London: Edward Arnold; 1973

Selected Websites for Access to Genetic and Genomic Information

Cancer Genome Atlas NIH. https://cancergenome.nih.gov/. Accessed June 6, 2017

Cancer Genome Projects ICGC. http://icgc.org/. Available at: http://www.icgc.org and http://www.sanger.ac.uk/perl/genetics/CGP/cosmic. Accessed January 24, 2012

Deciphering human disease. Database of causes of illness. Available at: http://www.sanger.ac.uk/. Accessed June 6, 2017

Encyclopedia of DNA Elements. ENCODE. Available at: http://www.genome.gov/encode. Accessed June 6, 2017

GeneTests, a clinical information resource relating genetic testing to the diagnosis, management, and genetic counseling of individuals and families with specific inherited disorders. At: https://www.genetests.org. Accessed June 6, 2017

Genome Bioinformatics UCSC Genome Browser. Available at: http://genome.ucsc.edu/. Accessed June 6, 2017

Genome-wide Association Studies. Available at: http://www.genome.gov/GWAStudies/. Accessed June 6, 2017

Human Epigenome Project (HEP). Available at: http://www.epigenome.org/. Accessed June 11, 2017

International HapMap Project. Available at: www.hapmap.ncbi.nlm.nih.gov. Accessed June 6, 2017

National Human Genome Research Institute. Available at: http://www.genome.gov/Planning/. Accessed June 6, 2017

MITOMAP. A human mitochondrial genome database. Available at: http://www.gen.emory.edu/mitomap.html. Accessed June 6, 2017

National Center for Health Statistics at Centers for Disease Control and Prevention. Available at: http://www.cdc.gov/nchs/. Accessed June 6, 2017

Nature Web Focus: Human Genome Collection. Available at: http://www.nature.com/nature/supple-ments/collections/humangenome/. Accessed January 24, 2012

OMIM. Online Mendelian Inheritance of Man. Available at: http://www.ncbi.nlm.nih.gov/omim. Accessed June 6, 2017

Online Metabolic and Molecular Bases of Inherited Disease. Available at: www.ommbid.com/. Accessed June 6, 2017

Thousand Genomes Project. Available at: http://www.1000genomes.org. Accessed June 6, 2017

Important Advances that Contributed to the Development of Genetics

(This list represents a selection and should not be considered complete; apologies to all authors not included.)

1665 Cells described and named (*Robert Hooke)*

1827 Human egg cell described *(Karl Ernst von Baer)*

1839 Cells recognized as the basis of living organisms *(Schleiden, Schwann)*

1859 Concept and facts of evolution *(Charles Darwin)*

1866 Rules of inheritance by distinct "factors" acting dominantly or recessively (*Gregor Mendel)*

1869 "Nuclein": a new acidic, phosphorus-containing, long molecule (F. *Miescher)*

1874 Monozygotic and dizygotic twins distinguished *(C. Dareste)*

1876 "Nature and nurture" (*F. Galton)*

1879 Chromosomes in mitosis *(W. Flemming)*

1883 Quantitative aspects of heredity (*F. Galton)*

1888 Term "chromosome" *(W. Waldeyer)*

1889 Term "nucleic acid" (*R. Altmann)*

1892 Term "virus" (*R. Ivanowski)*

1897 Enzymes discovered (*F. Biichner)*

1900 Mendel's discovery recognized *(H. de Vries, E. Tschermak, K. Correns,* independently). ABO blood group system *(Landsteiner)*

1901 Term "Mutation" coined *(H. deVries)*

1902 Some diseases in man inherited according to Mendelian rules *(W. Bateson, A. Garrod). Sex chromosomes (McClung). Chromosomes and Mendel's factors are related (W. Sutton). Individuality of chromosomes (T. Boveri)*

1906 Term "genetics" proposed *(W. Bateson)*

1907 Amphibian spinal cord culture *(Harrison)*

1908 Population genetics (*G. H. Hardy, W. Weinberg)*

1909 Inborn errors of metabolism (*A. Garrod). Terms "gene," "genotype," and "phenotype" proposed (W. Johannsen). Chiasma formation during meiosis (Janssens) First inbred mouse strain, DBA (C. Little)*

1910 Beginning of *Drosophila* genetics (*T. H. Morgan). First Drosophila mutation (white eyed)*

1911 Sarcoma virus (*Peyton Rous)*

1912 Crossing over (*T. H. Morgan and E. Cattell). Genetic linkage (T. H. Morgan and C. J. Lynch). First genetic map (A. H. Sturtevant)*

1913 First long-term cell culture (*A. Carrel). Nondisjunction (C. B. Bridges)*

1915 Genes located on chromosomes (chromosomal theory of inheritance; *Morgan, Sturtevant, Muller, Bridges). Bithorax mutant (C. B. Bridges). First genetic linkage in vertebrates (J. B. S. Haldane, A. D. Sprunt, N. M. Haldane). Term "intersex" (R. B. Goldschmidt)*

1917 Bacteriophage discovered (*F. d'Herrelle)*

1922 Characteristic phenotypes of different trisomies in the plant. *Datura stramonium* (*F. Blakeslee)*

1923 Chromosome translocation in *Drosophila* (*C. B. Bridges)*

1924 Blood group genetics (*Bernstein). Statistical analysis of genetic traits (R. A. Fisher)*

1926 Enzymes are proteins (*J. Sumner)*

1927 Mutations induced by X-rays (*H. J. Muller). Genetic drift (S. Wright)*

1928 Euchromatin/heterochromatin (*E. Heitz). Genetic transformation in bacteria (F. Griffith)*

1933 Pedigree analysis (*Haldane, Hogben, Fisher, Lenz, Bernstein). Polytene chromosomes (Heitz and Bauer, Painter)*

1934 Term "aneuploidy" coined *(A. F Blakeslee)*

1935 First cytogenetic map in *Drosophila (C B. Bridges). First estimate of human mutation rate (JBS Haldane)*

1937 Mouse H2 gene locus *(P. Gorer).* First human linkage group hemophilia A-colorblindness *(J. Bell and J. B. S. Haldane)*

1938 Telomere defined *(H J. Mutter)*

1940 Polymorphism (*E. B. Ford*). *Rhesus blood groups (Landsteiner and Wiener)*

1941 Evolution through gene duplication (*E. B. Lewis*). *Genetic control of enzymatic biochemical reactions (Beadle and Tatum). Mutations induced by mustard gas (C Auerbach and M. Robson)*

1942 Concept of epigenetics *(C. H. Waddington)*

1943 Mutations in bacteria *(S. E. Luria and M. Delbrück)*

1944 DNA as the material basis of genetic information *(Avery, MacLeod, McCarty). What is Life? The Physical Aspect of the Living Cell.* An influential book (*E. Schrödinger*)

1946 Genetic recombination in bacteria *(Lederberg and Tatum)*

1947 Genetic recombination in viruses *(Delbrück and Bailey, Hershey)*

1949 Sickle cell anemia, a genetically determined molecular disease *(Neel, Pauling).* Hemoglobin disorders prevalent in areas of malaria *(J. B. S. Haldane).* X chromatin *(Barr and Bertram)*

1950 Defined relation of the four nucleotide bases (*E. Chargaff*)

1951 Mobile genetic elements in Indian corn, *Zea mays (B. McClintock).* α-helix and β-sheet in proteins (l. *Pauling and R. B. Corey*)

1952 Genes consist of DNA *(Hershey and Chase).* Plasmids *(Lederberg).* Transduction by phages *(Zinder and Lederberg).* First enzyme defect in man (*Cori and Cori). First linkage group in man (Mohr). Colchicine and hypotonic treatment in chromosomal analysis (Hsu and Pomerat). Exogenous factors as a cause of congenital malformations (J. Warkany)*

1953 DNA structure (*Watson and Crick, Franklin, Wilkins). Conjugation in bacteria (W. Hayes, L. L. Cavalli, J. and E. Lederberg, independently). Non-Mendelian inheritance (Ephrussi). Cell cycle (Howard and Pelc). Dietary treatment of phenylketonuria (Bickel)*

1954 DNA repair (*Muller). HLA system (J. Dausset). Leukocyte drumsticks (Davidson and Smith). Cells in Turner syndrome are X-chromatin negative (P. Polani). Cholesterol biosynthesis (K. Bloch)*

1955 First genetic map at the molecular level (*S. Benzer). First amino acid sequence of a protein, insulin (F. Sanger). Lysosomes (C. de Duve). Buccal smear (Moore, Barr, Marberger). 5-Bromouracil, an analogue of thymine, induces mutations in phages (A. Pardee and R. Litman)*

1956 46 Chromosomes in man (*Tijo and Levan, Ford and Hamerton). Amino acid sequence of hemoglobin molecule (V. Ingram). DNA synthesis in vitro (S. Ochoa, A. Kornberg). Synaptonemal complex, the area of synapse in meiosis (M. J. Moses, D. Fawcett). Genetic heterogeneity (H. Harris, C. F. Fraser)*

1957 Genetic complementation (*Fincham). Genetic analysis of radiation effects in man (Neel and Schull)*

1958 Semiconservative replication of DNA (*M. Meselson and F. W. Stahl). Somatic cell genetics (G. Pontecorvo). Ribosomes (Roberts, Dintzis). Cloning of single cells (Sanford, Puck)*

1959 First chromosomal aberrations in man: trisomy 21 (*Lejeune, Gautier, Turpin). Turner syndrome, 45, XO (C. E. Ford). Klinefelter syndrome: 47 XXY (Jacobs and Strong). DNA polymerase (A. Kornberg). Isoenzymes (Vesell, Markert). Pharmacogenetics (Motulsky, Vogel)*

1960 Phytohemagglutinin-stimulated lymphocyte cultures (*Nowell, Moorhead, Hungerford*). Philadelphia chromosome (*Nowell and Hungerford)*

1961 The genetic code is read in triplets (*Crick, Brenner, Barnett, Watts-Tobin). The genetic code determined (Nirenberg, Mathaei, Ochoa). X-chromosome inactivation (M. F. Lyon, confirmed by Beutler, Russell, Ohno). Gene regulation, concept of operon (Jacob and Monod). Galactosemia in cell culture (Krooth). Cell hybrid-*

ization (Barski, Ephrussi). Thalidomide embryopathy (Lenz, McBride)

1962 Molecular characterization of immunoglobulins (*Edelman, Franklin). Identification of individual human chromosomes by ^{3}H-autoradiography (J. German, O. J. Miller). Term "codon" for a triplet of (sequential) bases (S. Brenner). Replicon (Jacob and Brenner). Cell culture (W. Szybalski and E. K. Szybalska). Xg, the first X-linked human blood group (Mann, Race, Sanger). Screening for phenylketonuria (Guthrie, Bickel)*

1963 Lysosomal storage diseases (*C. de Duve). First autosomal deletion syndrome (cri-du-chat syndrome, J. Lejeune)*

1964 Colinearity of gene and protein gene product (*C. Yanofsky). Excision repair (Setlow). MLC test (Bach and Hirschhorn, Bain and Lowenstein). Microlymphotoxicity test (Terasaki and McClelland). Selective cell culture medium HAT (J. Littlefield). Spontaneous chromosomal instability (J. German, T. M. Schröder). Cell culture from amniotic fluid cells (H. P. Klinger). Hereditary diseases studied in cell cultures (Danes, Bearn, Krooth, Mellman). Population cytogenetics (Court Brown). Fetal chromosomal aberrations in spontaneous abortions (Carr, Benirschke)*

1965 Sequence of alanine transfer RNA from yeast (*R. W. Holley). Limited life span of cultured fibroblasts (Hayflick, Moorhead). Crossing over in human somatic cells (J. German). Cell fusion with Sendai virus (H. Harris and J. F. Watkins)*

1966 Genetic code complete. Catalog of Mendelian phenotypes in man (*V. A. McKusick)*

1968 Restriction endonucleases (*H. O. Smith, Linn and Arber, Meselson and Yuan). Okazaki fragments in DNA synthesis (R. T. Okazaki). HLA-D the strongest histocompatibility system (Ceppellini, Amos). Repetitive DNA (Britten and Kohne). Biochemical basis of the ABO blood group substances (Watkins). DNA excision repair defect in xeroderma pigmentosum (Cleaver). First assignment of an autosomal gene locus in man (Donahue, McKusick). Synthesis of a gene in vitro (H. G. Khorana). Neutral gene theory of molecular evolution (M. Kimura)*

1970 Reverse transcriptase (*D. Baltimore, H. Temin, independently). Synteny, a new term to refer to all gene loci on the same chromosome (Renwick). Enzyme defects in lysosomal storage diseases (Neufeld, Dorfman). Individual chromosomal identification by specific banding stains (Zech, Casperson, Lubs, Drets and Shaw, Schnedl, Evans). Y chromatin (Pearson, Bobrow, Vosa). Thymus transplantation for immune deficiency (van Bekkum)*

1971 Two-hit theory in retinoblastoma (*A. G. Knudson)*

1972 High average heterozygosity (*Harris and Hopkinson, Lewontin). Association of HLA antigens and diseases*

1973 Receptor defects in the etiology of genetic defects, genetic hyperlipidemia (*Brown, Goldstein, Motulsky). Demonstration of sister chromatid exchanges with BrdU (S. A. Latt). Philadelphia chromosome as translocation (J. D. Rowley)*

1974 Chromatin structure, nucleosome (*Kornberg, Olins and Olins). Dual recognition of foreign antigen and HLA antigen by T lymphocytes (P. C. Doherty and R. M. Zinkernagel). Clone of a eukaryotic DNA segment mapped to a specific chromosome location (D. S. Hogness)*

1975 Southern blot hybridization (*E. Southern). Monoclonal antibodies (Köhler and Milstein). First protein signal sequence identified (G. Blobel). Model for promoter structure and function (D. Pribnow). First transgenic mouse (R. Jaenisch). Asilomar conference about recombinant DNA*

1976 Overlapping genes in phage ΦX174 (*Barell, Air, Hutchinson). Loci for structural genes on each human chromosome known (Baltimore Conference on Human Gene Mapping). First diagnosis using recombinant DNA technology (W. Kan, M. S. Golbus, A. M. Dozy)*

1977 Genes contain coding and noncoding DNA segments (*R. J. Roberts, P. A. Sharp,*

independently). First recombinant DNA molecule that contains mammalian DNA. Methods to sequence DNA (F. Sanger, Maxam and Gilbert). Sequence of phage ΦX174 (F. Sanger). X-ray diffraction analysis of nucleosomes (Finch and coworkers)

1978 Terms "exon" and "intron" for coding and noncoding parts of eukaryotic genes (*W. Gilbert*).
β-globulin gene structure (Leder, Weissmann, Tilghman and others). Mechanisms of transposition in bacteria. Production of somatostatin with recombinant DNA. Introduction of "chromosome walking" to find genes. First genetic diagnosis using restriction enzymes (Y. H. Kan and A. M. Dozy). DNA tandem repeats in telomeres (E. H. Blackburn and J. G. Gall)

1979 Small nuclear ribonucleoproteins ("snurps," *M R. Lerner and J. A. Steitz*). Alternative genetic code in mitochondrial DNA (*B G. Barell, A. T. Bankier, J. Drouin*). *p53 protein (D. P. Lane, A. Levine, L. Crawford, L. Old)*

1980 Restriction fragment length polymorphism for mapping *(D. Botstein and coworkers).* Genes for embryonic development in *Drosophila* studied by mutational screen (*C. Nüsslein-Volhard and E. Wieschaus*). *First transgenic mice by injection of cloned DNA (J. W. Gordon). Transformation of cultured mammalian cells by injection of DNA (M. R. Capecchi). Structure of 16S ribosomal ribonucleoprotein (C. Woese)*

1981 Sequence of a mitochondrial genome (*S. Anderson, S. G Barrett, A. T. Bankier)*

1982 Tumor suppressor genes *(H. P. Klinger).* Prions (proteinaceous infectious particles) as cause of central nervous system diseases (kuru, scrapie, Creutzfeldt-Jakob disease; *S. B. Prusiner). Insulin made by recombinant DNA marketed (Eli Lilly)*

1983 Cellular oncogenes *(H. E. Varmus and others).* HIV virus (*I. Montagnier, R. Gallo). Molecular basis of chronic myelocytic leukemia (C. R. Bartram, D. Bootsma and coworkers). First recombinant RNA molecule (E. A. Miele, D. R. Mills, F. R. Kramer). Bithorax complex of Drosophila sequenced (W. Bender)*

1984 Identification of the T cell receptor (*Tonegawa) Homeobox (Hox) genes in Drosophila and mice (W. McGinnis). Localization of the gene for Huntington disease (Gusella). Description of Helicobacter pylori (B. Marshall and R. Warren)*

1985 Polymerase chain reaction (*K. B. Mullis, R. K. Saiki). Hypervariable DNA segments as "genetic fingerprints" (A. Jeffreys). Hemophilia A gene cloned (J. Gietschier). Sequencing of the HIV-1 virus linkage analysis of the gene for cystic fibrosis (H. Eiberg and others). Isolation of telomerase from Tetrahymena (C. W. Greider and E. H. Blackburn). Isolation of a zinc finger protein from Xenopus oocytes (J. R. Miller, A. D. McLachlin, A. Klug). Insertion of DNA by homologous recombination (O. Smithies). Genomic imprinting in the mouse (B. Cattanach)*

1986 First cloning of human genes. Human visual pigment genes characterized (*J. Nathans, D. Thomas, D. S. Hogness). RNA as catalytic enzyme (T. Cech). First identification of a human gene based on its chromosomal location (positional cloning) (B. Royer-Pokora and coworkers)*

1987 Fine structure of an HLA molecule (*Björkman, Strominger and coworkers). Knockout mouse (M. Capecchi). A genetic map of the human genome (H. Donis-Keller and coworkers). Mitochondrial DNA and human evolution (R. L. Cann, M. Stoneking, A. C. Wilson)*

1988 Start of the Human Genome Project. Molecular structure of telomeres at the ends of chromosomes (*E. H. Blackburn and others). Cloning of the gene for Duchenne muscular dystrophy (L. M. Kunkel and others). Mutations in human mitochondrial DNA (D. C. Wallace). Transposable DNA as rare cause of hemophilia A (H. H. Kazazian). Successful gene therapy in vitro*

1989 Identification of the gene causing cystic fibrosis (*L. C. Tsui and others). Microdissection and cloning of a defined region of a human chromosome (Lüdecke, Senger, Claussen, Horsthemke)*

1990 Mutations in the *p53* gene as cause of Li–Fraumeni syndrome (*D. Malkin). Mutations in the gene wrinkled seed used by Mendel (M. K. Bhattacharyya). A defective gene as cause of inherited breast cancer (Mary-Claire King)*

1991 Odorant receptor multigene family (*Buck and Axel). Complete sequence of a yeast chromosome. Increasing use of microsatellites as polymorphic DNA markers. Trinucleotide repeat expansion as a new class of human pathogenic mutations*

1992 High-density map of DNA markers on human chromosomes. X chromosome inactivation center identified. p53 knockout mouse (*O. Smithies)*

1993 Gene for Huntington's disease cloned (*M. E. MacDonald). Developmental mutations in zebra fish (M. C. Mullins and C. Nüsslein-Volhard)*

1994 First physical map of the human genome in high resolution. Mutations in fibroblast growth factor receptor genes as cause of achondroplasia and other human diseases (*M. Muenke). Identification of genes for hereditary breast cancer*

1995 Cloning of the *BLM* (Bloom syndrome) gene (*N. A. Ellis, J. Groden, J. German and coworkers). First genome sequence of a free-living bacterium, Haemophilus influenzae (R. D. Fleischmann, J. C. Venter and coworkers). Master gene of the vertebrate eye, sey (small eye; G. Halder, P. Callaerts, W. J. Gehring). STS map of the human genome (T. J. Hudson and coworkers)*

1996 Yeast genome sequenced (*A. Goffeau and coworkers). Mouse genome map with more than 7,000 markers (E. S. Lander)*

1997 Sequence of *E. coli* (*F. R. Blattner and coworkers). Helicobacter pylori (J. F. Tomb). Neanderthal mitochondrial DNA sequences (M. Krings, S. Pääbo and coworkers). Mammal ("Dolly, the sheep") cloned by transfer of an adult cell nucleus into an enucleated oocyte (I. Wilmut)*

1998 RNA interference (RNAi; *A. Fire and coworkers). Nematode Caenorhabditis elegans genome sequenced human embryonic stem cells (Thomson and Gearhart)*

1999 First human chromosome (22) sequenced. Ribosome crystal structure

2000 *Drosophila* genome sequenced (*M. D. Adams). First complete genome sequence of a plant pathogen (Xylella fastidiosa). Arabidopsis thaliana, the first plant genome sequenced*

2001 First draft of the complete sequence of the human genome (*F. H. Collins, J. C. Venter and coworkers)*

2002 Genome sequence of the mouse (*R. H. Waterston and coworkers). Sequence of the genome of rice, Oryza sativa (J. Yu, S. A. Goff and coworkers). Sequence of the genomes of malaria parasite, Plasmodium falciparum and its vector, Anopheles gambiae. Earliest hominid, Sahelanthropos tchadensis (M. Brunet)*

2003 International HapMap Project and ENCODE launched sequence of the human Y chromosome (*H. Skaletsky, D. C. Page and coworkers). Homo sapiens idaltù, the oldest anatomically modern man from Pleistocene 154–160 years ago (T. D. White and coworkers)*

2004 Genome sequence of the Brown Norway rat. A new small-bodied hominin from Flores island, Indonesia *(P. Brown and coworkers)*

2005 Massive parallel DNA sequencing methods ("next generation sequencing") introduced genome sequence of the chimpanzee (*R.H. Waterston, E. S. Lander, R. K. Watson and coworkers). A total of 1.58 million human single-nucleotide polymorphisms mapped (D. A. Hinds, D. R Cox and coworkers). Human haplotype map. Sequence of the human X chromosome (M. T. Ross and coworkers). Inactivation profile of the human X chromosome (L Carrel and H. F. Willard)*

2006 All human chromosomes sequenced. Induced pluripotent stem cells (iPS), Takahashi & Yamanaka (Nobel 2012)

2007 Genome-wide studies applied to find predisposing factors for certain diseases. Genomic disorders recognized

2008 Synthetic bacterial genome (*C. Venter and coworkers.)*
Sequencing of individual human genomes

2009 Whole genome analysis by microarrays cancer genomes sequenced. *Ardipithecus ramidus* defines new stages in human evolution *(T. White and others)*

2010 Exome Sequencing Neanderthal genome sequence. Induced pluripotent stem (iPS) cells

2011 Genome structural variation (*E. E. Eichler and coworkers).*
Chromothripsis, a catastrophic event in oncogenesis (PJ Stephens and coworkers).

2012 Whole genome sequencing. Cancer epigenetics. Genome of a Denisovan.

Topologically associated domains (TADs) in chromatin.

2013 ENCODE project (**En**cyclopedia **O**f **D**NA **E**lements) launched. Genome editing CRISPR-Cas (*M. Jinek and coworkers, Charpentier & Doudna).*

2014 Nucleosome remodeling, SWI/SNF complex. Hallmarks of aging (*López-Otin).* Next generation sequencing. Complete Neandertal sequence (*K. Prüfer and coworkers). Lung cancer genome landscape.*

2015 Epigenome roadmap. *Homo naledi* (*L. Berger*). 1,000 Genomes Project. *Cancer Genome Atlas.*

2016 New telomere lengthening mechanism (*Dilley and coworkers). Mitochondrial replacement.*

2017 (first half) Liquid biopsy for ctDNA (circulating tumor DNA)

References for the Chronology

In addition to personal notes, dates are based on the following main sources:

Dunn LC. A Short History of Genetics. New York: McGraw-Hill; 1965

King RC, Mulligan PK, Stansfield WD. A Dictionary of Genetics. 8th ed. Oxford: Oxford University Press; 2013

Lander ES, Weinberg RA. Genomics: journey to the center of biology. Science 2000;287(5459):1777–1782 PubMed

McKusick VA. Human genetics: the last 35 years, the present, and the future. Am J Hum Genet 1992;50(4):663–670

McKusick VA, Harper PS. History of Medical Genetics. In: Rimoin D, Pyeritz R, Korf B, eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York: Elsevier; 2013. Chap 1

Watson JD. The Double Helix: A Personal Account of the Discovery of the Structure of DNA. Stent GS, ed. London: Weidenfeld & Nicolson; 1981

Sturtevant AH. A History of Genetics. New York: Cold Spring Harbor Press; 2001

The New Encyclopaedia Britannica. 15th ed. Chicago: Encyclopaedia Britannica; 2010

Vogel F, Motulsky AG. Human Genetics: Problems and Approaches. 3rd ed. Heidelberg: Springer-Verlag; 1997

Whitehouse HLK. Towards an Understanding of the Mechanism of Heredity. 3rd ed. London: Edward Arnold; 1973

Fundamentals

Phylogenetic Tree of Living Organisms

A phylogenetic tree attempts to show inferred evolutionary relationships of living organisms. The first example of such a tree was presented by Lamarck in 1809. It is the only figure in the *Origin of Species*, in which Charles Darwin in 1859 wrote, "Probably all of organic beings which have ever lived on this Earth have descended from some primordial form." There is an overall agreement that the earth is a little more than 4.5 billion years old and that early forms of life date back to approximately 3.5 billion years.

A. Three primary branches of the tree of life

The formal evolutionary hierarchy of groups of organisms proceeds from the largest to the smallest groups: domain → kingdom → phylum → order → class → family → genus → species. Living organisms are grouped according to the type of cells they consist of, either *prokaryotic* cells or *eukaryotic* cells. A third group of living organisms was recognized in the late 1960s, the Archaea (also called *archaebacteria*). They are assigned to two classes: Euryarchaeota and Crenarchaeota. Recent data indicate that the three-domain evolutionary tree actually consists of two domains: archaea and bacteria. Eukaryotes arose through a partnership between them, approximately 1.8 billion years ago (Williams et al., 2013).

Archaea can sustain without molecular oxygen at high (70–110°C, *thermophiles*) or low temperatures (*psychrophiles*), in water with high concentrations of sodium chloride (*halophiles*) or sulfur (*sulfothermophiles*), a highly alkaline environment (pH 11.5, alkaliphiles), in acid conditions (acidophiles), or in a combination of such adverse conditions that would boil or dissolve ordinary bacteria. Eukaryotes consist of several kingdoms, including animals, fungi, plants, algae, protozoa, and others. All domains have a presumed common progenitor called the *last universal common ancestor*.

B. Phylogeny of metazoa (animals)

The phylogeny of metazoa differs depending on whether it is based on the traditional interpretation or on molecular evidence as revealed mainly by ribosomal RNA sequence comparisons.

C. Mammalian phylogeny

Mammals arose approximately 100 to 120 million years ago in the late Mesozoic period of the earth. The timescale is only approximate. Of the 4,629 known mammalian species, 4,356 are placentals, and these fall into 12 orders. The first five placental orders according to their number of species are rodents (2015), followed by bats (925), insectivores (385), carnivores (271), and primates (233). DNA sequence data have resulted in some rearrangements of the phylogeny. (Figures adapted from Klein & Takahara, 2001)

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Allers T, et al. Archaeal genetics - the third way. Nat Rev Genet 2005; 6(1): 58–73

Delsuc F, et al. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 2005; 6(5): 361–375

Delsuc F, et al. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 2006; 439(7079): 965–968

Telford MJ. The animal tree of life. Science 2013; 339: 764–766

Klein J, Takahara N. Where Do We Come From? The Molecular Evidence for Human Descent. Berlin, Heidelberg: Springer-Verlag, 2001

Murphy WJ, et al. Molecular phylogenetics and the origins of placental mammals. Nature 2001; 409 (6820): 614–618

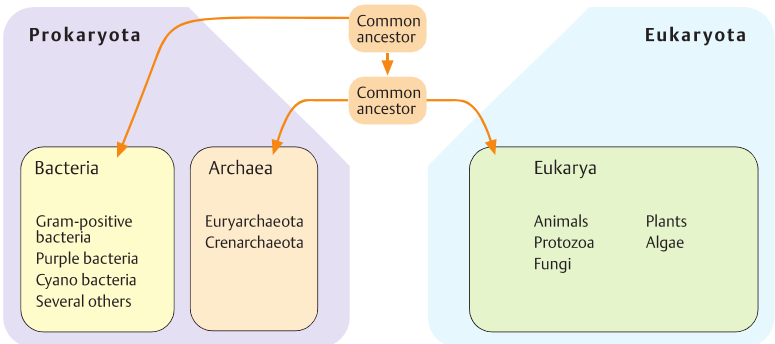
Rivera MC, Lake JA. The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature 2004; 431(7005): 152–155

Williams TA, et al. An archaeal origin of eukaryotes supports only two primary domains of life. Nature 2013; 504: 231–236

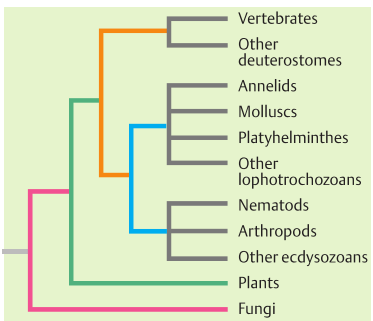
Woese CR. Interpreting the universal phylogenetic tree. Proc Natl Acad Sci U S A 2000; 97(15): 8392–8396

Woese CR. On the evolution of cells. Proc Natl Acad Sci U S A 2002; 99(13): 8742–8747

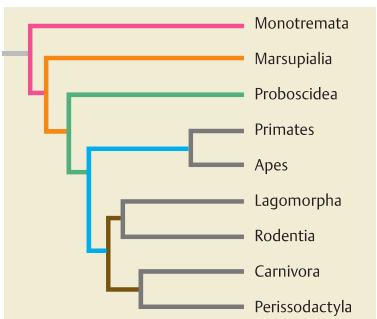
Woese CR. A new biology for a new century. Microbiol Mol Biol Rev 2004; 68(2): 173–186



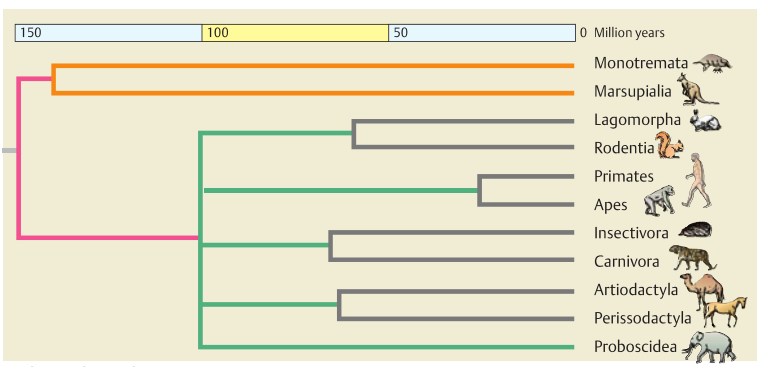
A. Three domains of living organisms (simplified version)



B. Phylogeny of metazoa, simplified



2. The molecular interpretation



1. The traditional view

C. Mammalian phylogeny (simplified)

Origins of Humans

Human and chimpanzee lineages separated from a common ancestor approximately 6 to 7 million years ago (mya). Several extinct species related to human evolution are collectively referred to as *hominids*, although the term *hominin* also is used. A complex pattern of human evolution has emerged during the past 10 to 15 years based on new fossils and genetic data derived from sequence analysis of ancient DNA. Important sites of hominids in Africa are along the Rift Valley in East Africa in the Afar and Hadar regions (north-East Ethiopia), the region on both sides of Lake Turkana in Kenya, Olduvai Gorge and Laetoli (Tanzania), two sites in Central Africa (Bar-el-Ghazal and Toros-Menalla, both in Chad), and several sites in South Africa (Sterkfontein, Kromdraai, Swartkrans, and Taung).

A. Time chart of hominid origins

Four hominid evolutionary age-related phases are distinguished: archaic hominids, transient forms including robust hominids, premodern humans, and anatomically modern man (AMM). Each group consists of different members with a relationship indicated by an arrow, however, it is often not known in detail. Of the early stages, only parts of a skeleton or teeth are available in most cases.

Archaic hominids: the oldest member of this group is *Sahelanthropus tchadensis* (6–7 mya), found in Central Africa, 2,500 km west of the Rift Valley. The size of the brain is that of a chimpanzee (360–370 cm^3), but the face is relatively flat, and enamel thickness is intermediate between a human and a chimpanzee. Two genera from approximately 5 to 6 mya are *Orrorin tugenensis* ("original man from the Tugen hills," Baringo region of central Kenya) and *Ardipithecus ramidus kadabba* (from Middle Awash, Afar, Ethiopia).

Transient hominids: this group comprises a subfamily of Hominidae, *Australopithecinae*, with a possible ancestral role for the early hominin species. Bipedal gait with concomitant anatomical changes of hands and feet, reduction of tooth size, and progressive development of the brain are main characteristics of this group. The member of this group, discovered first by Dart in 1924 (Taung child) in South Africa, was *Australopithecus africanus* with an apparent ability to walk upright. Best known of this group is *A. afarensis* (3–4 mya) with upright gait and other hominid features. "Lucy" is its most prominent representative. The robust hominids with a large jaw and teeth became extinct and are not considered to be ancestral to the genus *Homo*.

Premodern Man: the major representative of premodern man is *Homo erectus*. It originated in Africa approximately 1.9 mya and is the first earliest hominid to be found outside Africa in Asia and the Caucasus (1.6–1 mya in Georgia). *Homo habilis* (2.4–1 mya), "handy man," was not anagenetic to *H. erectus* but overlapped in time. The origin of the genus *Homo* coincides with the distinctive use of stone tools. Early *H. erectus* specimens from Africa are sometimes referred to as *H. ergaster* (2.3–1.4 mya). *Homo heidelbergensis* (0.6–0.1 mya) is an extinct *Homo* species that may be an ancestor of both *H. neanderthalensis* and *H. sapiens* in Europe. *Homo antecessor* (0.8 mya) is an extinct human species discovered 10 years ago in Gran Dolina, Spain. *Homo floresiensis* is a 60 mya extinct species with small stature (1 m adult height) discovered on Flora island, Indonesia (for Denisovans, see next page).

AMM: all humans living today belong to one species, *Homo sapiens*. *Homo sapiens* is derived from *H. erectus* in Africa approximately 200–100,000 years ago. AMM left Africa approximately 50,000 years ago and migrated to all continents at different times. (Data based on Wood, 2005; and Stringer, 2005.)

Further Reading

Brumm A, et al. Age and context of the oldest known hominin fossils from Flores. Nature 2016; 534 (7606): 249–253

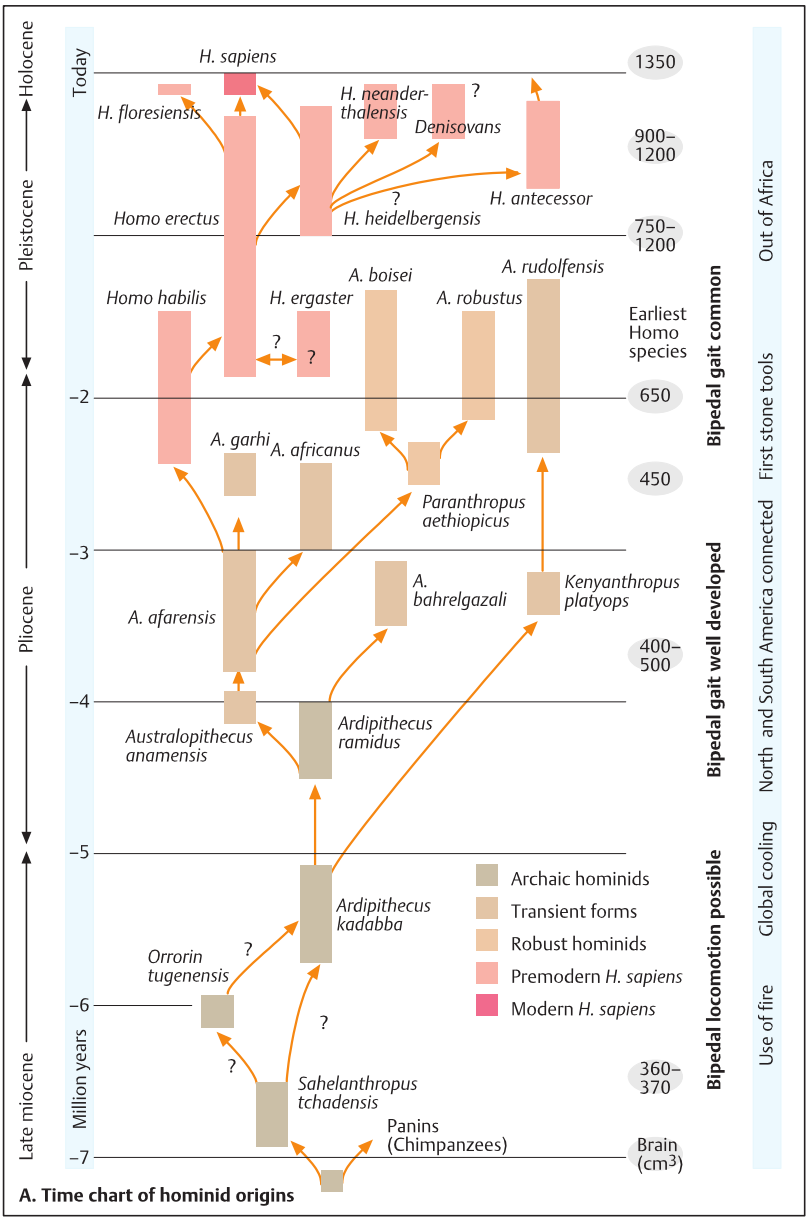
Jobling MA, et al. Human Evolutionary Genetics. 2nd ed. New York, NY: Garland, 2014

Kuhlwilm M, et al. Ancient gene flow from early modern humans into Eastern Neanderthals. Nature 2016; 530(7591): 429–433

Slatkin M, et al. Ancient DNA and human history. Proc Natl Acad Sci U S A 2016; 113(23): 6380–6387

Stringer C. The Origin of Our Species. London: Penguin Books, 2012

Wood B. Human Evolution. Oxford: Oxford University Press, 2005



A. Time chart of hominid origins

Out of Africa: Toward Modern Humans

Modern humans evolved in Africa approximately 0.2 mya. From there, several groups migrated out of Africa into Asia. DNA of all non-African human populations living today has been linked recently to one exodus. Modern humans arriving in Europe, approximately 50,000 years ago, met an older population that had been living there for approximately 0.4 my, the Neandertals. For the most part, they occupied different sites and developed different, albeit, overlapping forms of early culture, referred to as *Mousterian* for the Neandertals and *Acheulian* for the early humans, named after sites in France. In Asia, modern humans met an additional archaic hominin group, Denisovans, approximately 100 to 60,000 years ago.

A. Origins of humans in relation to Neandertals

Modern humans and Neandertals have shared a common ancestor approximately 0.8 mya. The population split occurred 270,000 to 440,000 years ago. For approximately 14,000 years, Neandertals and modern humans coexisted in regions extending from South-East Europe to Siberia. Even though Neandertals became extinct approximately 28,000 years ago, approximately 4% of European and Asian genomes share nuclear DNA with Neandertals, but African populations do not. A third group, named *Denisovans*, has been discovered from the Denisovan cave in the Altai Mountains, Siberia. Neandertals and Denisovans are more closely related to each other than to modern humans (Prüfer et al., 2014). Approximately 0.5% of the Denisovan genome was contributed by Neanderthals. (Figure adapted from Nonnan et al, 2006).

B. Dispersal of modern humans out of Africa

For the exodus out of Africa, approximately 60,000 years ago, northern and southern routes have been recognized. Asia was reached first (50,000–70,000 years ago), then Australia, and Europe somewhat later (40,000 years ago). The Americas were reached last (~13,000 years ago). The dispersal from Africa coincides with the growth of local populations and scarcity of resources. The climate must have been a main influence on the human populations that arrived in Europe. The use of tools developed rapidly, but cultural artifacts without practical use were also fabricated. The oldest tools for producing clothing, such as bone needles, are approximately 40,000 years old. Agriculture and domesticated animals were introduced approximately 12,000 years ago. (Figure modified from Jones, 2007).

C. Tools and art of early modern humans

Modern humans in the Upper Paleolithic age (17–33,000 years ago) were sophisticated makers of tools and art. Four stages of development have been identified: Magdalenian, Solutrean, Gravettian, and Aurignacian/Châtelperronian. The examples show a 26,000-year-old bone needle (**1**) and 38,000-year-old flutes made out of swan bone (**2**). (Photographs Conard et al, 2009).

Medical relevance

The field of evolutionary medicine deals with scientific and practical questions concerning the evolutionary background of many diseases in modern humans. Examples of such disorders are obesity, arteriosclerosis, hypertension, coronary heart disease, autoimmune diseases, and others. (Gluckman et al., 2009.)

Further Reading

Conard NJ, et al. New flutes document the earliest musical tradition in southwestern Germany. Nature 2009; 460(7256): 737–740

Gluckman P, et al. Principles of Evolutionary Medicine. Oxford: Oxford University Press, 2009

Jones D. Going global: how humans conquered the world. New Scientist, October 24, 2007

Meyer M, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. Nature 2016; 531(7595): 504–507

Noonan JP. Neanderthal genomics and the evolution of modern humans. Genome Res 2010; 20(5): 547–553

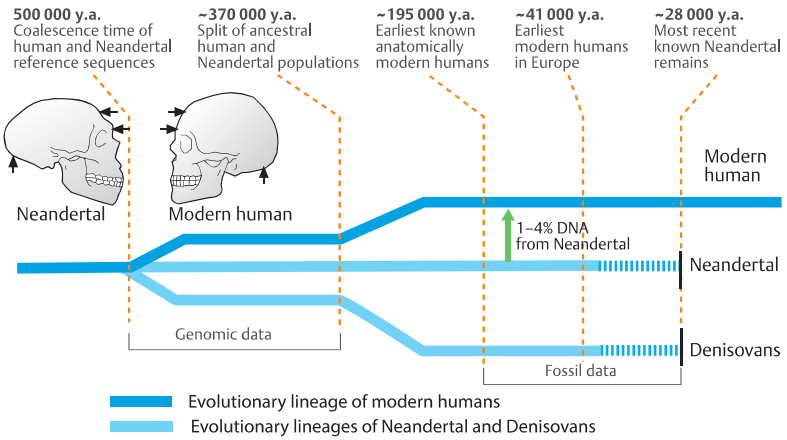
Noonan JP, et al. Sequencing and analysis of Neanderthal genomic DNA. Science 2006; 314(5802): 1113–1118

Prüfer K, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 2014; 505(7481): 43–49

Reich D, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 2010; 468(7327): 1053–1060

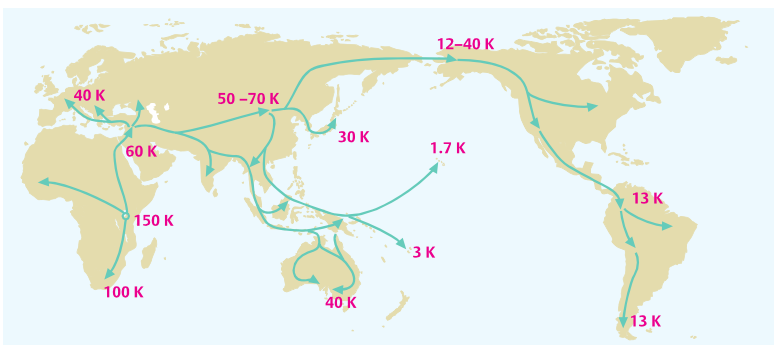
Roebroeks W, Soressi M. Neandertals revisited. Proc Nat Acad Sci 2016; 113: 6372–6379

Tucci S, et al. Population genetics: a map of human wanderlust. Nature 2016; 538(7624): 179–180

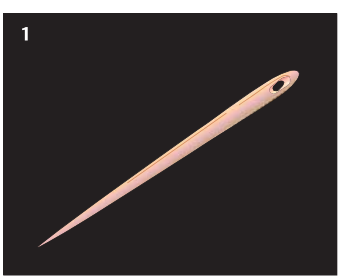


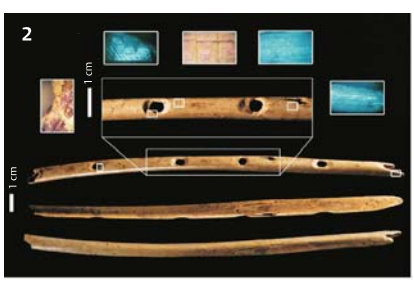
A. Origins of humans in relation to Neandertals

The branch lengths are schematic and not to scale. (y.a.: years ago)



B. Dispersal of modern humans out of Africa





C. Tools and art of early modern humans

The Cell and Its Components

Cells are the smallest organized structural units of living organisms, surrounded by a membrane, they are able to carry out a wide variety of functions during a limited life span. Each cell originates from another living cell, as postulated by Rudolf Virchow in 1855 (omnis celtula e cellula). Three basic types of cells exist: (1) *prokaryotic cells*, which carry their functional information in a circular genome without a nucleus, (2) the bacteria (or eubacteria) and archaea (or archaebacteria), and (3) *eukaryotic cells*, which contain their genome in individual chromosomes in a nucleus. They have a well-organized internal structure. Robert Hooke introduced the word *cell* in 1665 for the tiny cavities in cork, as they reminded him of the small rooms in which monks sleep. Cells were recognized as the "elementary particles of organisms," animals, and plants by Mathias Schleiden and Theodor Schwann in 1839. Today, we understand most of the biological processes in cells at the molecular level.

A. Scheme of a prokaryotic cell

Prokaryotic cells (bacteria) are typically rod-shaped or spherical, a few micrometers in diameter, and without a nucleus or special internal structures. Within a cell wall consisting of a bilayered cell membrane, bacteria contain on average 1,000 to 5,000 genes tightly packed in a circular molecule of DNA. In addition, they usually contain small circular DNA molecules named *plasmids*. These *plasmids* replicate independently of the main chromosome and generally contain genes that confer antibiotic resistance.

B. Scheme of a eukaryotic cell

A eukaryotic cell consists of cytoplasm and a nucleus, enclosed by a plasma membrane. The eukaryotic cell nucleus contains the genetic information. Nucleoli are nonmembrane-bound sites of ribosome synthesis. The nuclear membrane separates the nucleus from the cytoplasm, which harbors a complex system of inner membranes, that form discrete structures (organelles). These are the *mitochondria* (in which important energy-delivering chemical reactions take place), the *endoplasmic reticulum* (a series of membranes in which important molecules are formed), the *Golgi apparatus* (for transport functions), *lysosomes* (in which some proteins are broken down), and *peroxisomes* (for the formation or degradation of certain molecules).

Animal cells (**1**) and plant cells (**2**) share several features, but differ in important structures. A plant cell contains chloroplasts for photosynthesis, surrounded by a rigid wall of cellulose and other polymeric molecules, and contains vacuoles for water, ions, sugar, nitrogen-containing compounds, or waste products. Vacuoles are permeable to water but not to the other substances enclosed within them.

C. Plasma membrane of the cell

Cells are surrounded by a plasma membrane composed of bipartite molecules of fatty acids, water-repellent phospholipids arranged in a double layer (bilayer). Numerous molecules traverse the plasma membrane for special functions in cell communication. Different types of membrane proteins can be distinguished: (1) transmembrane proteins used as channels to transport molecules into or out of the cell, (2) proteins connected with each other to provide stability, (3) receptor molecules involved in signal transduction, (4) molecules with enzyme function to catalyze internal chemical reactions in response to an external signal, and (5) gap junctions in specialized cells forming pores between adjacent cells. Gap junction proteins are composed of connexins. They allow the passage of molecules as large as 1.2 nm in diameter. Cells contain four major families of organic molecules: carbohydrates, fatty acids, amino acids, and nucleotides.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

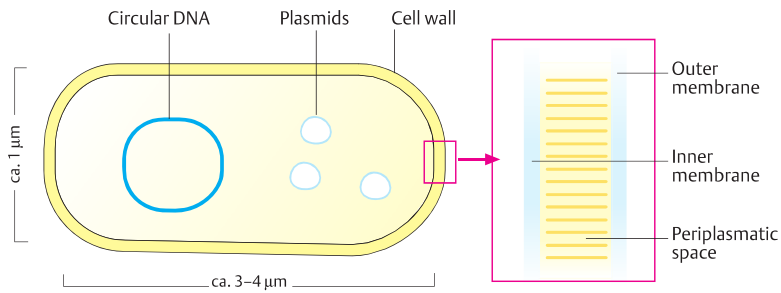
Challoner J. The Cell: A Visual Tour of the building Block of Life. Chicago, IL: University of Chicago Press, 2015

de Duve C. A Guided Tour of the Living Cell. Vols. 1 and 2. New York, NY: Scientific American Books, 1984

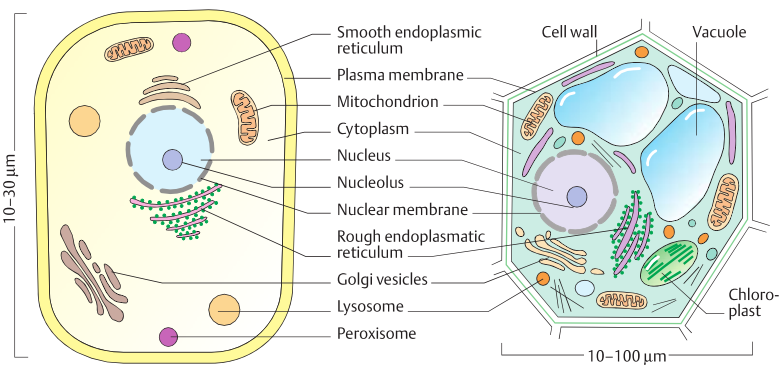
The Human Protein Atlas. Available at http://www.proteinatlas.org. Accessed 30 August, 2017

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

The Human Cell. Poster in Science. November 25, 2016. Available at http://science.sciencemag.org/content/354/6315/1055.3. Accessed 30 August, 2017

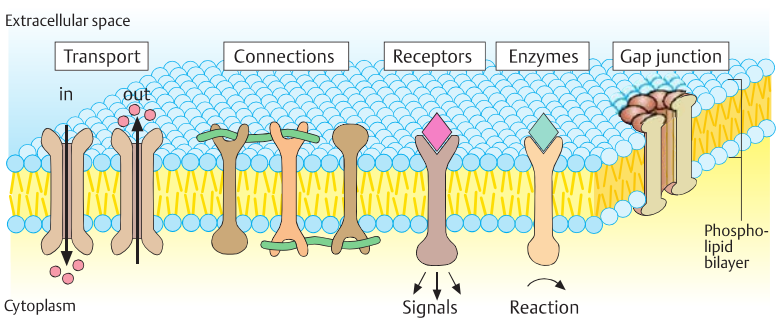


A. Scheme of a prokaryotic cell



1. Animal cell **2.** Plant cell

B. Scheme of an eukaryotic cell



C. Plasma membrane

Genetic Background of Aging Processes

Aging is the result of time-dependent processes affecting cellular, genetic, and metabolic homeostasis, including stem cell regenerative capacity. Organisms age within a species-specific *life span*, which suggests that aging not only involves damage but also genetically programmed factors. Certain gene mutations can extend life span in nematodes, Drosophila, and mice. Some human genetic disorders are associated with premature aging and a reduced life span.

A. Phenotypic manifestations of aging

Gustav Klimt (1862–1918) illustrated how human phenotypes appear at different ages (**1**). At the cellular level, the life span of cultured human fibroblasts is limited to approximately 40 to 60 cell divisions, depending on the age of the donor, referred to as Hayflick phenomenon. Initially (phase I) cells grow as long as there is sufficient surface (phase II). Then they reach phase III and no longer divide (**2**). This quiescent, nondividing state is called *cellular senescence*. In contrast, cancer cells in culture continue to divide. At the organismal level, aging affects almost all functions of the body, cognitive abilities, and reproductive success. It also increases the risk for common diseases such as atherosclerosis and cancer. Seven pillars of areas of research in aging involving different mechanisms in aging processes have been suggested (**3**). (Illustrations: 1, Galleria Nazionale d'Arte Moderna, Roma; 2, Hayflick and Moorhead, 1961; 3, Kennedy et al., 2014.)

B. Age-related changes in the cell

Aging is associated with DNA damage and telomere shortening. In addition, other cellular structures and processes are impaired, including nuclear architecture, mitochondrial function and stability, epigenetic regulation, maintenance of proteins, and intercellular communication. (Figure adapted from López-Otín et al, 2013).

C. Genetic premature aging disorders

Adult progeria or Werner syndrome, WS (OMIM 277700), first described in 1904, is an autosomal recessive disorder of multiple premature aging manifestations with onset during the second and third decades of life. WS is caused by homozygous mutations in the DNA helicase gene *WRN* (OMIM 604611) encoding RECQL2. Other helicase disorders are Bloom syndrome (OMIM 210900) due to mutations in RECQL3 (*BLM* gene 604610), and Rothmund-Thomson syndrome (OMIM 277700) due to mutations in RECQL4 (603780). Helicases unwind the DNA and are important for DNA replication, recombination, repair, and transcription (see p. 52). Loss of helicase function leads to genome instability. (Illustration: Werner Syndrome International Registry, http://www.wernersyndrome.org/).

Further Reading

Criscione SW, et al. The chromatin landscape of cellular senescence. Trends Genet 2016; 32(11): 751–761

Dong X, et al. Evidence for a limit to human lifespan. Nature 2016; 538(7624): 257–259

Hayflick L, et al. The serial cultivation of human diploid cell strains. Exp Cell Res 1961; 25: 585–621

Hayflick L. The limited in vitro lifetime of human diploid cell strains. Exp Cell Res 1965; 37: 614–636

Kennedy BK, et al. Geroscience: linking aging to chronic disease. Cell 2014; 159(4): 709–713

Kenyon CJ. The genetics of ageing. Nature 2010; 464 (7288): 504–512

López-Otín C, et al. The hallmarks of aging. Cell 2013; 153(6): 1194–1217

Martin GM. The biology of aging. In: Longo DL, et al., eds. Harrison's Principles and Practice of Internal Medicine. 18th ed. New York, NY: McGrawHill Medical, 2012: 562–569

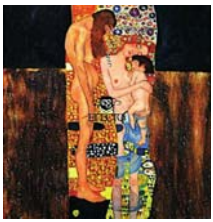
Oberdoerffer P, et al. The role of nuclear architecture in genomic instability and ageing. Nat Rev Mol Cell Biol 2007; 8(9): 692–702

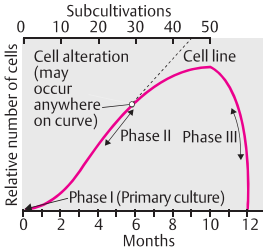
OMIM: Online Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders. Available at www.ncbi.nlm.nih.gov/Omim

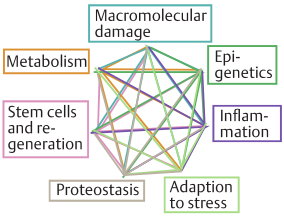
Oshima J, et al. Werner syndrome: clinical features, pathogenesis and potential therapeutic interventions. Aging Res Rev 2017; 33: 105–114

Zhang W, et al. Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. Science 2015; 348(6239): 1160–1163

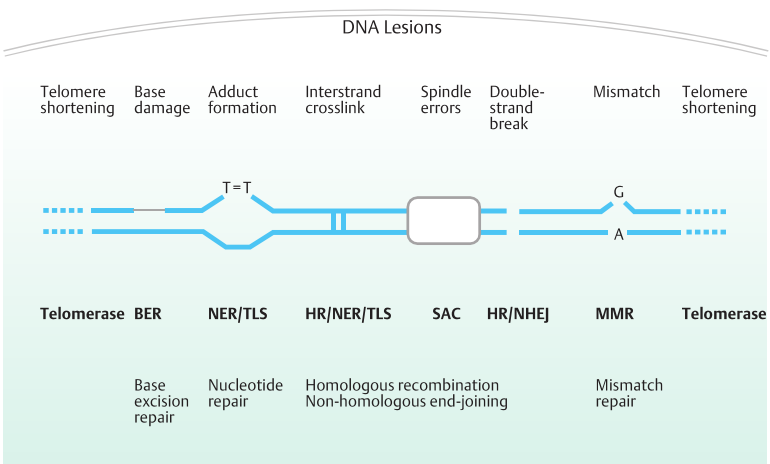
Zhang H, et al. NAD^+ repletion improves mitochondrial and stem cell function and enhances life span in mice. Science 2016; 352(6292): 1436–1443







A. Phenotypes of manifestations of aging



B. Age-related genomic changes

C. Werner syndrome – genetic premature aging

Carbohydrates

Carbohydrates are carbonyl compounds (aldehydes, ketones) that occur widely in living organisms as part of biomolecules. Carbohydrates are one of the most important classes of biomolecules. Their main functions can be classified into three groups: (1) to deliver and store energy, (2) to help provide the basic framework for DNA and RNA, the information-carrying molecules (see p. 46), and (3) to form structural elements of cell walls of bacteria and plants (polysaccharides). In addition, they form cell surface structures (receptors) used in conducting signals from cell to cell. Combined with numerous proteins and lipids, carbohydrates are important components of numerous internal cell structures.

A. Monosaccharides

Monosaccharides (simple sugars) are aldehydes (—C=O, —H) or ketones (C=O) with two or more hydroxyl groups (general structural formula: CH_2O). The aldehyde or ketone group can react with one of the hydroxyl groups to form a ring. This is the usual configuration of sugars that have five or six carbon atoms (pentoses and hexoses). The carbon atoms are numbered sequentially. The D (dextro) and L (levo) forms of sugars are mirror-image isomers of the same molecule.

The naturally occurring forms are the D forms. These further include β and α forms as stereoisomers. In the cyclic forms, the carbon atoms of sugars are not on a plane but they three-dimensionally take the shape of a chair or a boat. The arrangement of the hydroxyl groups can differ, and therefore stereoisomers such as mannose or galactose are formed.

B. Disaccharides

These are compounds of two monosaccharides linked with each other. Usually, a bond is formed between a hydroxyl group of one sugar and a hydroxyl group of another. Sucrose and lactose are frequently occurring disaccharides.

C. Derivatives of sugars

Glucuronic acid, glucosamine, and N-acetylglucosamine are sugar derivatives formed by replacing certain hydroxyl groups by other groups. They play a role in several genetic metabolic diseases.

D. Polysaccharides

Short (oligosaccharides) and long chains of sugars and sugar derivatives (polysaccharides) form essential structural elements of the cell. Complex oligosaccharides with bonds to proteins or lipids are part of cell surface structures, for example, blood group antigens. Others are part of the extracellular matrix of cells.

Medical relevance

Examples of human hereditary disorders in the metabolism of carbohydrates are as follows:

Diabetes mellitus (OMIM 125850): a heterogeneous group of disorders characterized by elevated levels of blood glucose, with complex clinical and genetic features (see p. 276).

Disorders of fructose metabolism: hereditary fructose intolerance (OMIM 229600), and hereditary fructose-1,6-bisphosphatase deficiency with hypoglycemia and often with lethal outcome in newborn infants (OMIM 229700).

Galactose metabolism: galactosemia (OMIM 230400), galactokinase deficiency (OMIM 230200), galactose epimerase deficiency (OMIM 230350), and others.

Glycogen storage diseases: there are eight types of disorders of glycogen metabolism that differ in clinical symptoms and the genes and enzymes involved (OMIM 232200, 232210–232800). In a large group of genetically determined disorders, complex polysaccharides cannot be degraded owing to reduced or absent enzyme function as storages diseases (mucopolysaccharidoses, mucolipidoses) (see p. 292).

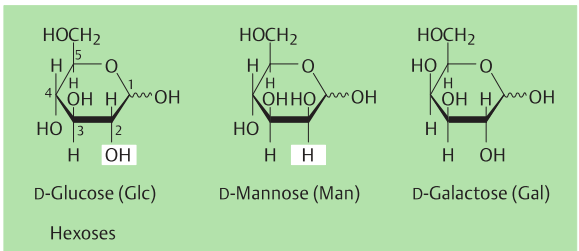
Further Reading

Koolman J, Roehm KH. Color Atlas of Biochemistry. 2nd ed. Stuttgart: Thieme, 2005

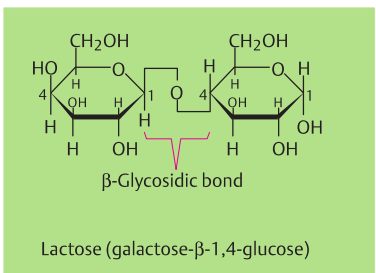
McKusick VA. Online Mendelian Inheritance in Man (OMIM). 12th ed. Baltimore, MD: Johns Hopkins University Press, 1998. Available at www.ncbi.nlm.nih.gov/Omim

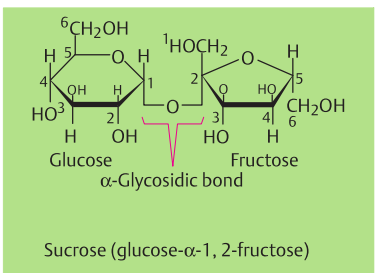
Ramachandran TS, ed. Disorders of Carbohydrate Metabolism. Available at http://emedicine.medscape.com/article/1183033-overview. Accessed November 1, 2016

Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease, OMMBD. Available at http://ommbid.mhmedical.com. Last printed version: Scriver CR, et al. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001

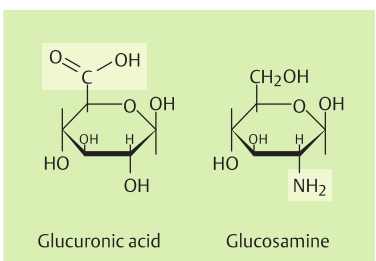


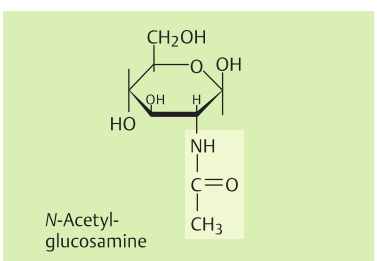
A. Monosaccharides



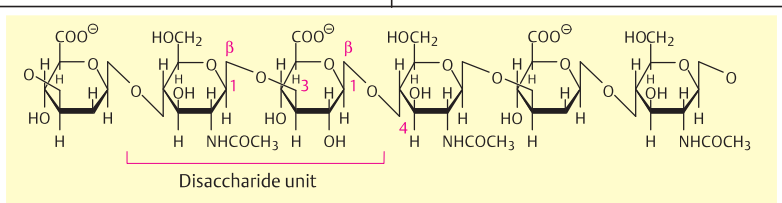


B. Disaccharides





C. Sugar derivatives



D. Polysaccharides

Lipids (Fatty Acids)

Fatty acids are essential components of cell membranes and precursors of biomolecules, such as steroid hormones and other molecules, used in signal transduction. Lipids are important energy-carrying components of food (dosage-dependent). They also form important compounds with carbohydrates (glycolipids) and phosphate groups (phospholipids). Lipids can be classified according to their ability to undergo hydrolytic cleavage (hydrolyzable and nonhydrolyzable).

A. Fatty acids

A fatty acid is composed of an unbranched hydrocarbon chain of 4 to 24 carbon atoms with a terminal carboxylic acid group (**1**). It is polar, with a hydrophilic head (—COOH) and a hydrophobic tail ($—CH_3$). Saturated fatty acids without a double bond and unsaturated fatty acids with one or more double bonds exist (**2**). Palmitic acid has a chemically very reactive hydrophilic carboxyl (—COOH) group at its head, and a hydrophobic tail that is not very reactive. Oleic acid is an unsaturated fatty acid with a double bond between carbon atoms 8 and 9. Linoleic acid and arachidonic acid are essential in human nutrition.

B. Lipids

Fatty acids can combine with other groups of molecules to form various types of lipids. As water-insoluble (hydrophobic) molecules, they are soluble in organic solvents only. The carboxyl group can enter into an ester or an amide bond. Triglycerides are compounds of fatty acids with glycerol.

Glycolipids (lipids with sugar residues) and phospholipids (lipids with a phosphate group attached to an alcohol derivative) are the structural bases of important macromolecules. Sphingolipids are an important group of molecules in biological membranes. Here, sphingosine, instead of glycerol, is the fatty-acid-binding molecule. Sphingomyelin and gangliosides contain sphingosine. Gangliosides make up 6% of the central nervous system lipids. They are degraded by a series of enzymes.

C. Lipid aggregates

Owing to their bipolar properties, fatty acids can form lipid aggregates in water. The hydrophilic ends are attracted to their aqueous surroundings; the hydrophobic ends protrude from the surface of the water and form a surface film. If completely under the surface, lipids may form a micelle, which is compact and dry within. Phospholipids and glycolipids can form two-layered membranes (lipid membrane bilayers). These are the basic structural elements of cell membranes.

D. Other lipids: steroids

Steroids are small molecules consisting of four different rings of carbon atoms. Cholesterol is the precursor of five major classes of steroid hormones: prostagens, glucocorticoids, mineralocorticoids, androgens, and estrogens. Each of these hormone classes is responsible for important biological functions, such as maintenance of pregnancy, fat and protein metabolism, blood volume and blood pressure, and sexual development.

Medical relevance

Several groups of disorders of lipoprotein and lipid metabolism exist. Important examples are familial hypercholesterolemia (OMIM 143890, see p. 284), hyperlipoproteinemia (OMIM 238600), dysbetalipoproteinemia (OMIM 107741), and high-density lipoprotein binding protein (OMIM 142695). Genetically determined disorders of ganglioside catabolism lead to severe diseases, for example, Tay–Sachs disease (OMIM 272800) caused by defective degradation of ganglioside GM2 (deficiency of β-*N*-acetylhexosaminidase), several types of gangliosidoses (OMIM 230500, 305650), Sandhoff disease (OMIM 268800), and others.

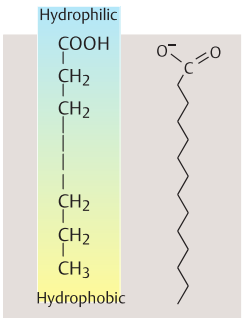
Further Reading

Gilbert-Barness E, Barness L. Metabolic Diseases: Foundations of Clinical Management, Genetics, and Pathology. Natick. MA: Eaton Publishing, 2000

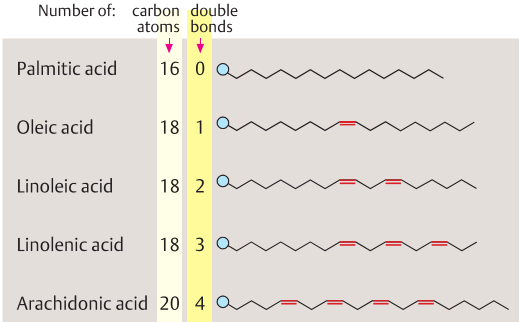
Koolman J, Roehm KH. Color Atlas of Biochemistry. 2nd ed. Stuttgart: Thieme, 2005

McKusick VA. Online Mendelian Inheritance in Man (OMIM). 12th ed. Baltimore, MD: Johns Hopkins University Press, 1998. Available at www.ncbi.nlm.nih.gov/

Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease, OMMBD. Available at http://ommbid.mhmedical.com. Last printed version: Scriver CR, et al. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001

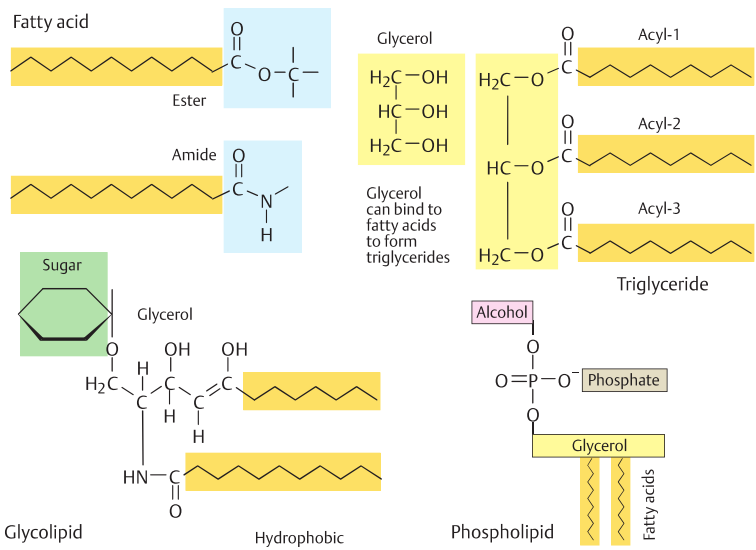


1. Saturated fatty acid

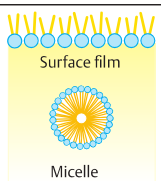


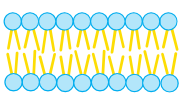
2. Common fatty acids

A. Fatty acids



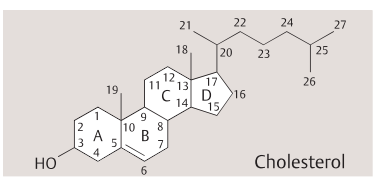
B. Lipids





Two-layered plasma membrane (phospholipid bilayer)

C. Lipid aggregates



D. Other lipids: steroids

Amino Acids

Amino acids (2-aminocarboxylic acids) are the basic structural units of proteins. An amino acid consists of a central α carbon atom that is bonded to four different chemical groups: one amino group (—NH_2), one carboxyl group (—COOH), a hydrogen atom (H), and a side chain (R). R is the major determinant of the individual functional property of each amino acid. At pH 7, both the NH_2 and COOH groups are ionized. Twenty different side chains confer different physicochemical properties to each amino acid. Amino acids exist in two mirror-image forms: D (dextro) and L (levo) isomers. Only the L forms occur in proteins, with rare exceptions. Amino acids are classified according to their side chains and chemical reactivity. For each amino acid, a three-letter and a one-letter abbreviation is used. Essential amino acids that cannot by synthesized by humans are Val, Leu, Ile, Phe, Trp, Met, Thr, and Lys. These have to be supplied by food intake.

A. Aliphatic amino acids

Amino acids with an aliphatic side chain are glycine with one hydrogen atom (—H) as a side chain, alanine with a methyl group (—CH_3), or a larger, hydrophobic (water-repellent) side chain as in valine, leucine, and isoleucine. Proline has a side chain bound to both the central carbon and the amino group in a ring structure. Hydrophobic aromatic side chains occur in phenylalanine (a phenyl group bound through a methylene [—CH_2—] group) and tryptophan (an indole ring bound through a methylene group). Two hydrophobic amino acids contain sulfur (S) atoms: cysteine with a sulfhydryl group (—SH) and methionine with a thioether (—S—CH_3). The sulfhydryl group in cysteine is very reactive and forms stabilizing disulfide bonds (—S—S—). These play an important role in stabilizing the three-dimensional forms of proteins. Selenocysteine is a cysteine analogue occurring in a few proteins, such as the enzyme glutathione peroxidase.

B. Hydrophilic amino acids

Serine, threonine, and tyrosine contain hydroxyl groups (—OH). Thus, they are hydrolyzed forms of glycine, alanine, and phenylalanine. The hydroxyl groups make them hydrophilic and reactive. Both asparagine and glutamine contain an amino group and an amide group. At physiological pH, their side chains are negatively charged.

C. Charged amino acids

These amino acids have either two ionized amino groups (basic) or two carboxyl groups (acidic). Basic amino acids (positively charged) are arginine, lysine, and histidine. Histidine has an imidazole ring, and it can be uncharged or positively charged, depending on its surroundings. It is frequently found in the reactive centers of proteins (e.g., in the oxygen-binding region of hemoglobin). Aspartic acid and glutamic acid have two carboxyl groups (—COOH) and are usually acidic. Seven of the 20 amino acids have slightly ionizable side chains, making them highly reactive (asparagine, glutamic acid, histidine, cysteine, tyrosine, lysine, and arginine).

Medical relevance

Glycine, phenylalanine, tyrosine, histidine, proline, and lysine, and the branched chain amino acids valine, leucine, and isoleucine are involved in genetic diseases showing toxic metabolic symptoms when their plasma concentration is too high or too low (disorders of amino acid metabolism).

Phenylketonuria: disorders of phenylalanine hydroxylation result in variable clinical signs and severity (OMIM 261600, see p. 278).

Maple syrup urine disease: a variable disorder caused by deficiency of branched chain α-keto acid dehydrogenase, which leads to accumulation of valine, leucine, and isoleucine (OMIM 248600). The classic severe form results in severe neurologic damage to the infant.

Further Reading

Gilbert-Barness E, Barness L. Metabolic Diseases: Foundations of Clinical Management, Genetics, and Pathology. Natick, MA: Eaton Publishing, 2000

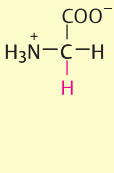
Koolman J, Roehm KH. Color Atlas of Biochemistry. 2nd ed. Stuttgart: Thieme, 2005

Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease, OMMBD. Available at http://ommbid.mhmedical.com. Last printed version: Scriver CR, et al. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001

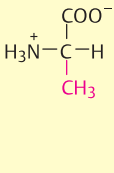
A. Neutral amino acids, nonpolar side chains

* Essential amino acids for humans

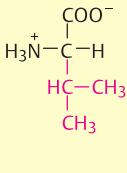
Aliphatic



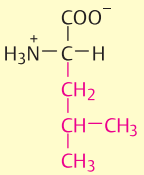
Glycine
Gly (G)



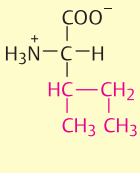
Alanine
Ala (A)



Valine*
Val (V)

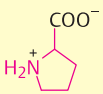


Leucine*
Leu (L)



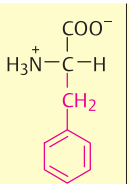
Isoleucine*
Ile (I)

Cyclic

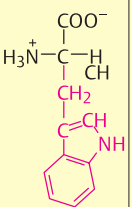


Proline
Pro (P)

Aromatic

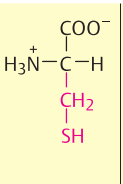


Phenylalanine*
Phe (F)

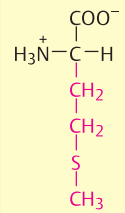


Tryptophan*
Trp (W)

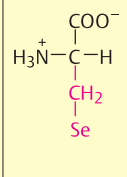
Sulfur-containing



Cysteine
Cys (C)



Methionine*
Met (M)

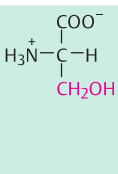


Seleno-
cysteine
Sec (U)

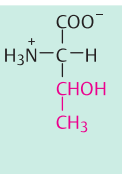
B. Hydrophilic amino acids, polar side chains

* Essential amino acids for humans

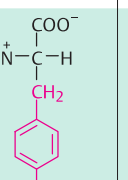
Aromatic



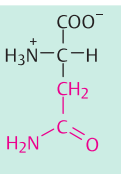
Serine
Ser (S)



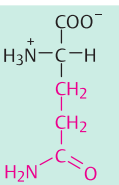
Threonine*
Thr (T)



Tyrosine Tyr (Y)



Asparagine
Asn (N)

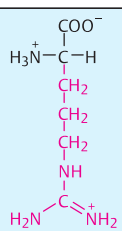


Glutamine
Gln (Q)

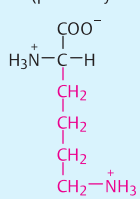
C. Charged amino acids

* Essential amino acids for humans

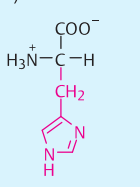
1. Basic (positively charged)



Arginine Arg (R)

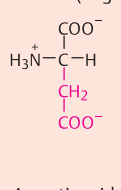


Lysine Lys (K)*

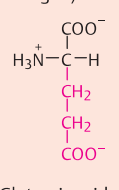


Histidine His (H)

2. Acid (negatively charged)



Aspartic acid
Asp (D)



Glutamic acid
Glu (E)

Nucleotides and Nucleic Acids

Nucleic acids are macromolecules that as subunits of DNA and RNA are central to the storage and transmission of genetic information. They participate in numerous biological processes, convey energy, are part of essential coenzymes, and regulate numerous metabolic functions.

A. Phosphate groups

Phosphate groups occur in nucleic acids and nucleotides as monophosphates (one phosphorus atom), diphosphates (two phosphorus atoms), or triphosphates (three phosphorus atoms).

B. Sugar residues

The carbohydrate residues in nucleotides are usually derived from either ribose as β-D-ribose (in RNA) or β-D-deoxyribose (in DNA).

C. Nucleotide bases of pyrimidine

Cytosine (C), thymine (T), and uracil (U) are the three pyrimidine nucleotide bases. They differ from each other in their side chains (—NH_2 on the carbon atom in position 4 [C-4] of cytosine, —CH_3 on C-5 in thymine, and O on C-4 in uracil). In addition, cytosine has a double bond between N-3 and C-4.

D. Nucleotide bases of purine

Adenine (A) and guanine (G) are the two nucleotide bases of purine. They differ in their side chains and in having a double bond between N-1 and C-6 (present in adenine, absent in guanine).

E. Nucleosides and nucleotides

A *nucleoside* is a compound of a sugar residue (ribose or deoxyribose) and a nucleotide base. The bond is between the carbon atom in position 1 (C-1) of the sugar and a nitrogen atom of the base (*N*-glycosidic bond). A *nucleotide* is a compound of a five-carbon-atom sugar residue (ribose or deoxyribose) attached to a nucleotide base (pyrimidine or purine base) and a phosphate group.

The nucleosides of the various bases are grouped as ribonucleosides or deoxyribonucleosides, for example, adenosine or deoxyadenosine, guanosine or deoxyguanosine, cytidine or deoxycytidine. Uridine occurs only as a ribonucleoside. Thymidine occurs only as a deoxyribonucleoside.

F. Nucleic acid

A nucleic acid consists of a series of nucleotides. A phosphodiester bridge between the 3′ carbon atom of one nucleotide and the 5′ carbon atom of the next nucleotide joins two nucleotides. The linear sequence is usually given in the 5′ to 3′ direction with the abbreviations of the respective nucleotide bases. For instance, ATCG would signify the sequence adenine (A), thymine (T), cytosine (C), and guanine (G) in the 5′ to 3′ direction.

Medical relevance

The following diseases are examples of human hereditary disorders in purine and pyrimidine metabolism:

Hyperuricemia and gout: a group of disorders resulting from genetically determined excessive synthesis of purine precursors (OMIM 240000).

Lesch–Nyhan syndrome: a variable, usually severe infantile X-chromosomal disease with marked neurologic manifestations resulting from hypoxanthine–guanine phosphoribosyltransferase deficiency (OMIM 308000).

Adenosine deaminase deficiency: a heterogeneous group of disorders resulting in severe infantile immunodeficiency. Different autosomal recessive and X-chromosomal types exist (OMIM 102700).

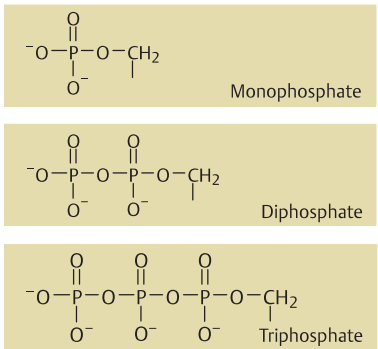
Further Reading

Gilbert-Barness E, Barness L. Metabolic Diseases: Foundations of Clinical Management, Genetics, and Pathology. Natick, MA: Eaton Publishing, 2000

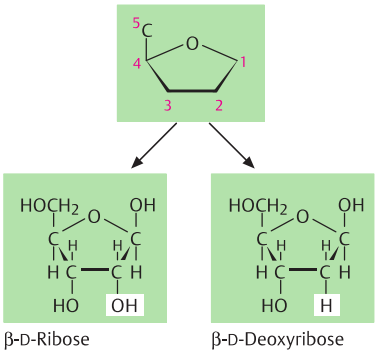
McKusick VA. Online Mendelian Inheritance in Man (OMIM). 12th ed. Baltimore, MD: Johns Hopkins University Press, 1998. Available at www.ncbi.nlm.nih.gov/OMIM

Koolman J, Roehm KH. Color Atlas of Biochemistry. 2nd ed. Stuttgart: Thieme, 2005

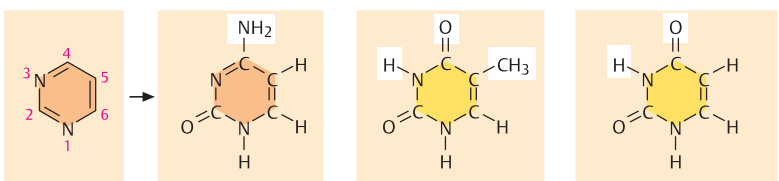
Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease, OMMBD. Available at http://ommbid.mhmedical.com. Last printed version: Scriver CR, et al. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001



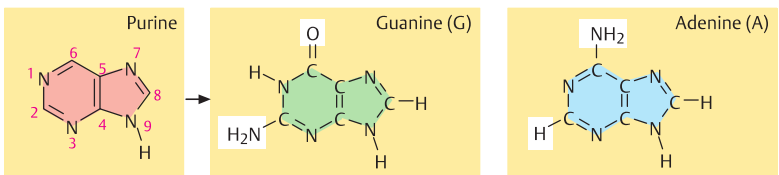
A. Phosphate groups



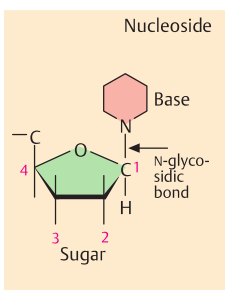
B. Sugar residues (pentose)

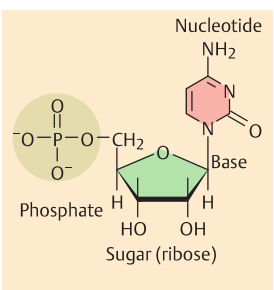


C. Nucleotide bases of pyrimidine

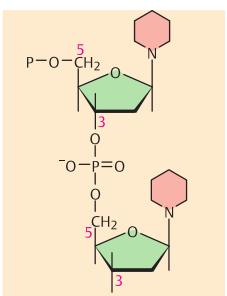


D. Nucleotide bases of purine





E. Nucleosides and nucleotides



F. Nucleic acid

DNA and Its Components

DNA consists of three components of DNA: nucleotide bases, phosphate groups, and residues of a sugar, deoxyribose. The components are arranged in a chain of nucleotides. They form a three-dimensional structure from which DNA derives its functional consequences (see p. 56). A second type of nucleic acid contains ribose (ribonucleic acid [RNA]).

A. Nucleotide bases

The nucleotide bases in DNA are heterocyclic molecules derived from either *purine* or *pyrimidine*. Five bases occur in the two types of nucleic acids DNA and RNA. The purine bases are adenine (A) and guanine (G). The pyrimidine bases are thymine (T) and cytosine (C) in DNA; in RNA, thymine is replaced by uracil (U). The nucleotide bases are part of a subunit of DNA, the nucleotide. A nucleotide consists of all three components of DNA: one of the four nucleotide bases, a sugar (deoxyribose), and a phosphate group. The nitrogen atom in position 9 of a purine or in position 1 of a pyrimidine is bound to the carbon in position 1 of the sugar (N-glycosidic bond). RNA differs from DNA in two respects: it contains ribose instead of deoxyribose (unlike the latter, ribose has a hydroxyl group on the carbon atom at position 2), and uracil instead of thymine. Uracil does not have a methyl group on the carbon atom at position 5. A nucleoside consists of a nucleotide base linked to position 1 of a pentose sugar.

B. DNA nucleotide chain

DNA is a linear polymer of deoxyribonucleotide units. The nucleotide chain is formed by joining a hydroxyl group on the sugar of one nucleotide to the phosphate group attached to the sugar of the next nucleotide. The sugars linked together by the phosphate groups form the invariant part of the DNA. The variable part is in the sequence of the nucleotide bases A, T, C, and G. A DNA nucleotide chain is polar. The polarity results from the way the sugars are attached to each other. The phosphate group at position C-5 (the 5′ carbon) of one sugar joins to the hydroxyl group at position C-3 (the 3′ carbon) of the next sugar by means of a phosphodiester bridge. Thus, one end of the chain has a 5′ phosphate group free and the other end has a 3′ hydroxyl group free (5′ end and 3′ end, respectively). By convention, the sequence of nucleotide bases is written in the 5′ to 3′ direction.

C. Hydrogen bonds between bases

The chemical structure of the nucleotide bases determines a defined spatial relationship. A purine (adenine or guanine) always lies opposite to a pyrimidine (thymine or cytosine). Three hydrogen-bond bridges form between cytosine (C) and guanine (G). Two hydrogen bonds form between adenine (A) and thymine (T). Therefore, either guanine and cytosine, or adenine and thymine are posed opposite each other, forming complementary base pairs G-C and A-T. Other spatial relationships are not possible. The distance between two bases is 2.90 and 3.00 Å, respectively.

D. Double strand of DNA

DNA forms two strands in opposite direction (antiparallel): one 5′ to 3′ end, and the other in the 3′ to 5′ end direction. The sequence of the nucleotide bases on one strand of DNA (in the 5′ to 3′ direction) is complementary to the nucleotide base sequence (or simply the base sequence) of the other strand in the 3′ to 5′ direction. The specificity of base pairing is the most important structural characteristic of DNA.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Berg JM, et al. Biochemistry. 7th ed. New York, NY: W. H. Freeman, 2011

Koolman J, Roehm KH. Color Atlas of Biochemistry. 2nd ed. Stuttgart: Thieme, 2005

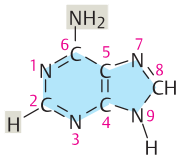
Krebs JE, et al. Kilpatrick: Lewin's Genes XI. Sudbury, MA: Jones & Bartlett, 2014

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

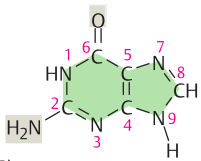
Olinski R, et al. Uracil in DNA–its biological significance. Mutat Res 2010; 705(3): 239–245

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Purines

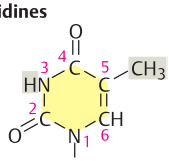


Adenine (A)



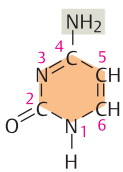
Guanine (G)

Pyrimidines



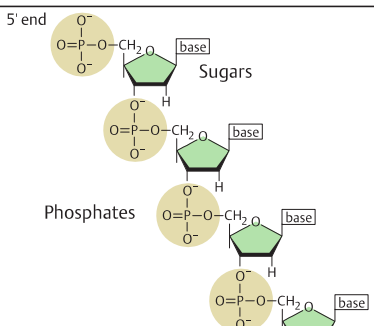
Thymine (T)

Uracil (U)



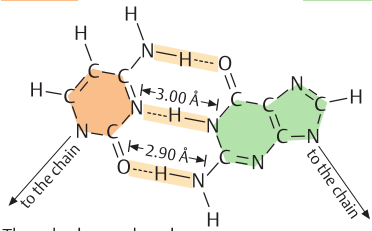
Cytosine (C)

A. Nucleotide bases



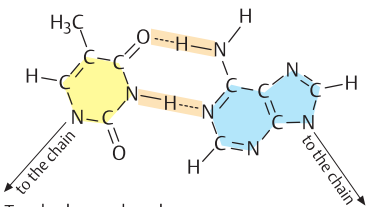
B. DNA nucleotide chain

Cytosine Guanine



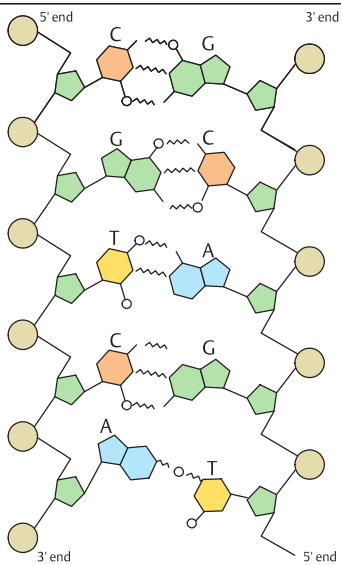
Three hydrogen bonds

Thymine Adenine



Two hydrogen bonds

C. Hydrogen bonds between bases



D. Double strand of DNA

DNA as a Carrier of Genetic Information

It took a surprisingly long time to realise that DNA stores and transmits genetic information, as opposed to proteins, which was assumed earlier. Friedrich Miescher's discovery in 1869 of "nuclein," a new, acidic, phosphorus-containing substance, later named *nucleic acid* by Richard Altmann in 1889, did not reveal its biological role. By 1900, the purine and pyrimidine bases were known, and 20 years later, the two kinds of nucleic acid, RNA and DNA, were distinguished. An incidental but precise observation in 1928 by Fred Griffith, and decisive experiments in 1944 by Oswald Avery, indicated that DNA is the carrier of genetic information.

A. Griffith's observation

In 1928, the English microbiologist Fred Griffith made a remarkable observation while investigating various strains of pneumococcus bacteria *(Streptococcus pneumoniae*, a cause of inflammation of the lungs, pneumonia). He determined that mice injected with strain S (smooth) died (**1**). On the other hand, animals injected with strain R (rough) survived (**2**). When he inactivated the lethal S strain by heat, the animals again survived (**3**). Surprisingly, a mixture of the nonlethal R strain and the heat-inactivated S strain had the same lethal effect as the original S strain (**4**). When he found normal living pneumococci of the S strain in the blood of the animals, he concluded that cells of the R strain must have changed into cells of the S strain. This is called *bacterial transformation*. For some time, this surprising result could not be explained and was met with skepticism. Its relevance for genetics was not apparent. (Figure adapted from Stent & Calendar, 1978).

B. The transforming principle is DNA

Griffith's findings formed the basis for investigations by Avery et al., in 1944 at the Rockefeller Institute in New York. They determined that the chemical basis of the transforming principle was DNA. From cultures of an S strain (**1**), they produced an extract of lysed cells (cell-free extract, **2**). After all the proteins, lipids, and polysaccharides had been removed, the extract retained the ability to transform pneumococci of the R strain into pneumococci of the S strain (transforming principle, **3**). Avery et al., determined that the cause was DNA alone. Thus, the DNA explained Griffith's observation. The DNA of the bacterial chromosomes remained intact after heating. The section of the chromosome with the gene responsible for capsule formation (S gene) could be released from the destroyed S cells and taken up by some R cells in subsequent cultures. After the S gene was incorporated into its DNA, an R cell was transformed into an S cell (**4**). This observation is based on the ability of bacteria to take up foreign DNA, which alters (transforms) some of their genetic attributes.

C. DNA transmits genetic information

The final evidence that DNA transmits genetic information was provided by Hershey and Chase (1952). They labeled the capsular protein of bacteriophages with radioactive sulfur (^{35}S) and the DNA with radioactive phosphorus (^{32}P). When bacteria were infected with the labeled bacteriophage, only ^{32}P (DNA) entered the cells and not ^{35}S (capsular protein). The subsequent formation of new, complete phage particles in the cell proved that DNA was the exclusive carrier of the genetic information needed to form new phage particles.

Further Reading

Avery OT, et al. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med 1944; 79(2): 137–158

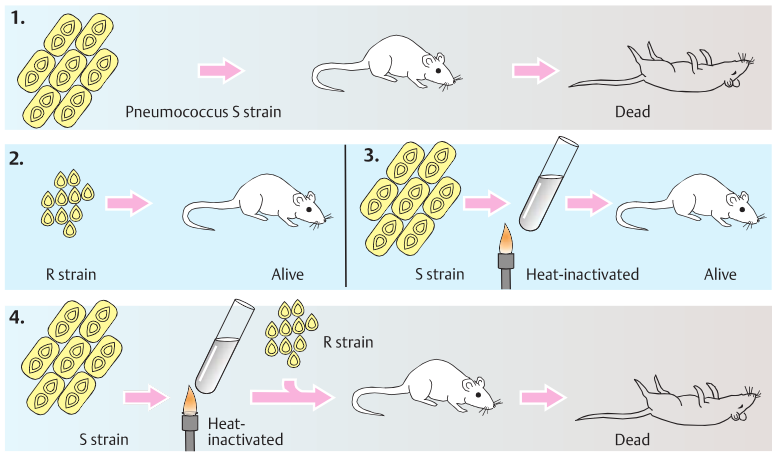
Griffith F. The significance of pneumococcal types. J Hyg (Lond) 1928; 27(2): 113–159

Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol 1952; 36(1): 39–56

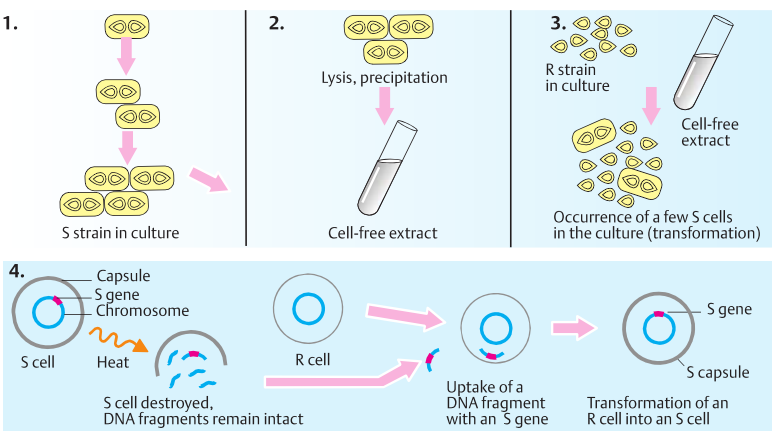
Judson MF. The Eighth Day of Creation. Makers of the Revolution in Biology. Expanded ed. New York, NY: Cold Spring Harbor Laboratory Press, 1996

McCarty M. The Transforming Principle. Discovering that Genes Are Made of DNA. London: W. W. Norton & Co, 1986

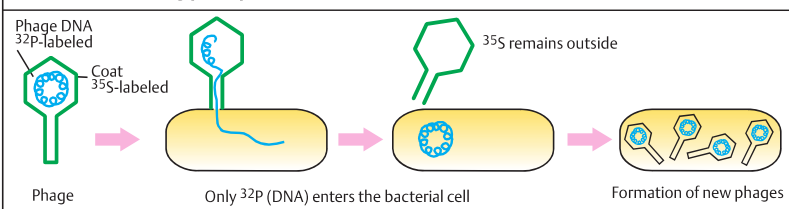
Stent GS, Calendar R. Molecular Genetics. An Introductory Narrative. 2nd ed. San Francisco, CA: W. H. Freeman, 1978



A. Griffith's observation



B. The transforming principle is DNA



C. Genetic information is transferred by DNA only

DNA Structure

The elucidation of the structure of DNA as a double helix in 1953 by Watson and Crick with important contributions by Rosalind Franklin and Maurice Wilkins is considered a landmark discovery; the cornerstone of modern genetics. The novel features of this structure explain two fundamental genetic mechanisms: storage of genetic information in a linear, readable pattern, and replication of genetic information to ensure its faithful transmission from generation to generation.

A. DNA double helix

The two complementary polynucleotide chains of DNA are wound around each other in opposite direction as a double helix along a common axis. The nucleotide bases lie inside the molecule in pairs, (an adenine opposite to a thymine, A–T, or a guanine opposite to a cytosine, G–C). The sugar–phosphate backbone lies on the outside. A minor and a major groove are distinguished. DNA can occur in alternative forms as A-DNA and Z-DNA. The most common form is B-DNA with approximately 10 base pairs per turn.

B. Replication

The complementary arrangement of the base pairs allows to make an exact copy (replicate) of a DNA molecule. Each strand serves as a template to generate a new strand after the double helix has been opened. DNA replication is semiconservative. This means that one completely new complementary strand is copied from each of the existing ones.

C. Denaturation and renaturation

DNA is stable at physiological temperatures because the two strands are held together by the many hydrogen bonds along the several hundred million nucleotides. The two strands separate by unwinding when exposed to increasing temperature or weak chemical reagents (e.g., alkali, formamide, or urea). This reversible process is called *denaturation* or melting. The melting temperature (T_m) depends on many factors, including the proportion of G–C pairs, because the three hydrogen bonds between G and C are more stable than the two bonds between A–T pairs. The resulting single-stranded DNA molecules do not maintain their helical structure. Lowering the temperature, increasing the ion concentration, or neutralizing the pH will reassociate the two strands into a double helix (*renaturation*), but only if they are complementary. If identical, they hybridize rapidly; if closely related, they hybridize slowly; if unrelated, they do not hybridize. The ability of single-stranded DNA to hybridize to another indicates that both strands are complementary. Hybridization of single-stranded DNA or RNA is widely used to explore whether DNA or RNA fragments contain complementary sequences.

D. Transmission of genetic information

The sequence of the nucleotide base pairs (A–T and G–C) is decoded and read in two steps called *transcription* and *translation*. In transcription, the sequence of one of the two DNA strands is converted into a complementary sequence of bases in a similar molecule, known as messenger RNA (mRNA). One DNA strand, the one in the 3′ to 5′ direction (coding strand), serves as the template. In translation, this sequence is converted into a defined sequence of amino acids. A triplet sequence, three base pairs (a codon), encodes one of the 20 amino acids. Beginning with a defined start point (methionine), the nucleotide sequence of the mRNA molecule is translated into a corresponding sequence of amino acids. DNA and RNA differ with respect to one nucleotide: RNA contains uracil (U) instead of the structurally related thymine. The mechanisms and biochemical reactions of transcription and translation are complex (not shown here).

Further Reading

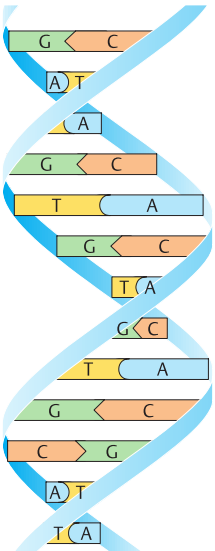
Watson JD, et al. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953; 171(4356): 737–738

Watson JD, et al. Genetical implications of the structure of deoxyribonucleic acid. Nature 1953; 171 (4361): 964–967

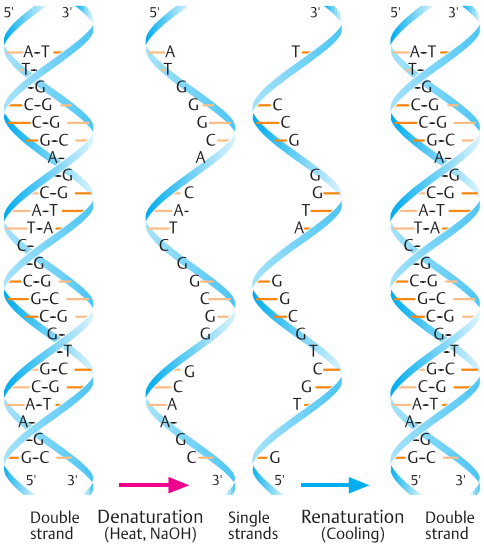
Wilkins MH, et al. Molecular structure of deoxypentose nucleic acids. Nature 1953; 171(4356): 738–740

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

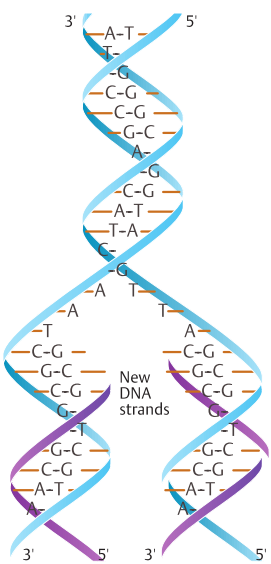
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016



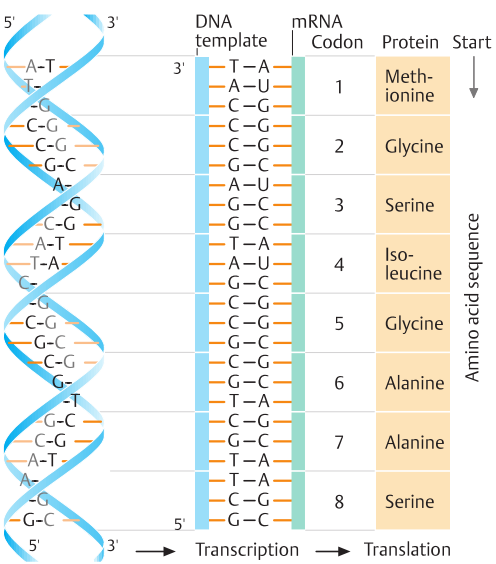
A. DNA double helix



C. Denaturation and renaturation



B. Replication



D. Transmission of genetic information

DNA Replication

DNA replication refers to the process of copying each DNA strand into a new complementary strand. This occurs at every cell division and ensures that genetic information is transmitted to both daughter cells. DNA replication requires the highly-coordinated action of many proteins in a complex called the *replisome*. Precision and speed are required. The two new DNA chains are assembled at a rate of approximately 1,000 nucleotides per second in *Escherichia coli*. During replication, each preexisting strand of DNA serves as a template for the formation of a new strand. This is referred to as semiconservative replication.

A. Prokaryotic replication begins at one site

In prokaryote cells, replication begins at a defined point in the ring-shaped bacterial chromosome, the origin of replication (**1**). From here, new DNA is formed in both directions. Replication can be visualized by autoradiography after the newly replicated DNA has incorporated tritium (^{3}H)-labeled thymidine (**2**).

B. Eukaryotic replication begins at several sites

In eukaryotic cells, DNA synthesis occurs during a defined phase of the cell cycle (S phase). Replication of eukaryotic DNA begins at numerous sites (replicons) (**1**). At each site, the parental strands are unwound by helicases. Replication proceeds in both directions from each replicon until neighboring replicons fuse (**2**) and the entire DNA is duplicated (**3**). The electron micrograph (**4**) shows replicons at three sites (arrows).

C. DNA replication fork

During DNA replication, a characteristic structure forms at the site of the opened double helix where the new strands are synthesized, the replication fork (**1**). This is a multienzyme complex containing enzymes for synthesis of DNA. It moves along the double helix as the parental strands are unwound by DNA helicases. Before this, an enzyme called *topoisomerase I* binds to DNA at random sites and relieves torsional stress by breaking a phosphodiester bond in one strand (making a nick). Each of the preexisting strands serves as a template for the synthesis of a new strand. However, the replication fork is asymmetric. One daughter DNA strand is synthesized continuously in the leading strand, whereas the other strand lags slightly behind (lagging strand). New DNA is synthesized in the 5′ to 3′ direction only because at the 3′ end, nucleotides can be attached continuously to the leading strand (**2**). At the 5′ end (the lagging strand), this is not possible.

Replication requires a primer of complementary RNA at the start site. At the leading strand, synthesis of new DNA proceeds continuously from a single primer in the 5′ to 3′ direction. In contrast, at the lagging strand, DNA is synthesized in small segments of 1,000 to 2,000 bases (Okazaki fragments) in the opposite direction. Each fragment requires its own primer. Subsequently, the primers are removed and replaced by DNA chain growth from the neighboring fragment, and the gaps are closed by DNA ligase. The enzyme responsible for DNA synthesis, DNA polymerase III, is complex and comprises several subunits. There are different enzymes for the leading and lagging strands in eukaryotes. A mismatch proofreading and repair system eliminates replication errors.

Further Reading

Alberts B. DNA replication and recombination. Nature 2003; 421(6921): 431–435

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Cairns J. The bacterial chromosome and its manner of replication as seen by autoradiography. J Mol Biol 1963; 6: 208–213

Chagin VO, et al. Organization of DNA replication. Cold Spring Harb Perspect Biol 2010; 2(4): a000737

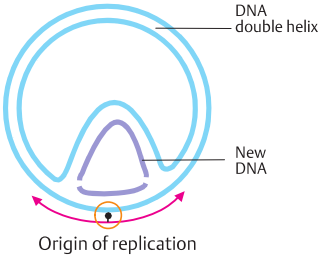
Kriegstein HJ, et al. Mechanism of DNA replication in Drosophila chromosomes: structure of replication forks and evidence for bidirectionality. Proc Natl Acad Sci U S A 1974; 71(1): 135–139

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

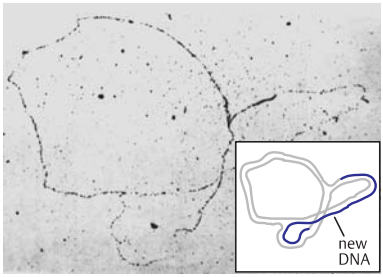
Marx J. How DNA replication originates. Science 1995; 270(5242): 1585–1587

Meselson M, Stahl FW. The replication of DNA in *Escherichia coli*. Proc Natl Acad Sci U S A 1958; 44(7): 671–682

Watson JD, et al. Molecular Biology of the Gene. 6th ed. New York, NY: Cold Spring Harbor Laboratory Press, 2008

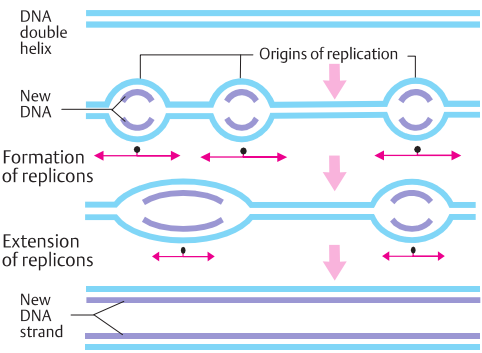


1. DNA replication in the bacterial chromosome

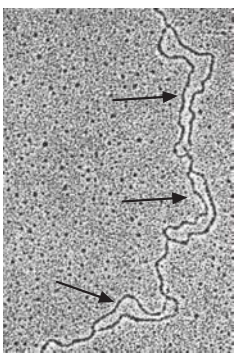


2. Prokaryotic replication in an autoradiogram in *E. coli* (J. Cairns)

A. Prokaryotic replication begins at one site

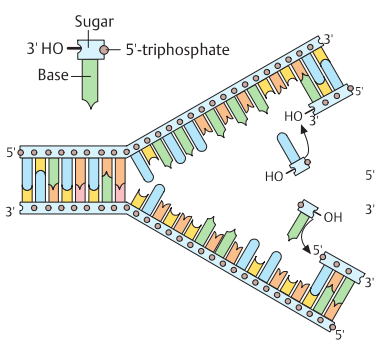


3. Replication completed

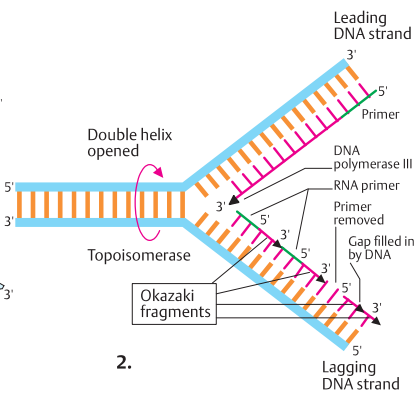


4. Eukaryotic replication in the EM (D. S. Hogness)

B. Eukaryotic replication begins at several sites



1.



2.

C. DNA replication fork

The Flow of Genetic Information: Transcription and Translation

The information contained in the nucleotide sequence of a gene is converted into useful biological function in two major steps: *transcription* and *translation*. This flow of genetic information is unidirectional. First, the information of the coding sequences of a gene is transcribed into an intermediary RNA molecule, which is synthesized in sequences that are precisely complementary to those of the coding strand of DNA (transcription). Second, the sequence information in the messenger RNA (mRNA) molecule is translated into a corresponding sequence of amino acids (translation). In eukaryotes, RNA is not used directly for translation. It is first processed into an mRNA molecule before it can be used. In prokaryotes, mRNA is directly transcribed from DNA.

A. Transcription

Here the nucleotide sequence of one strand of DNA is transcribed into a complementary molecule of RNA (mRNA). The DNA helix is opened by a complex set of proteins (different DNA helicases). The DNA strand in the 3′ to 5′ direction (coding strand) serves as the template for the transcription of DNA into RNA by RNA polymerase. RNA is synthesized in the 5′ to 3′ direction. This strand is the RNA sense strand. RNA transcribed under experimental conditions from the opposing DNA strand is called *antisense RNA*. It inhibits regular transcription.

B. Translation

Translation refers to the process by which the nucleotide sequence of the mRNA is used to construct a chain of amino acids (a polypeptide chain) in the sequence encoded in the DNA. Translation occurs in a reading frame, which is defined at the start of translation (start codon, AUG). Translation involves two further types of RNA molecule in addition to mRNA: transfer RNA (tRNA) and ribosomal RNA. tRNA deciphers the codons. Each amino acid has its own set of tRNAs, which bind the amino acid and carry it to the end of the growing polypeptide chain. Each tRNA has a region, called an *anticodon*, which is complementary to a codon of the mRNA. The figure shows codons 1, 2, 3, and 4 of the mRNA that have been translated into the amino acid sequence: methionine (Met), glycine (Gly), serine (Ser), and isoleucine (Ile). Glycine and alanine are added next in this example.

C. Stages of translation

Translation is accomplished in three steps. First, at initiation (**1**), an initiation complex comprising mRNA, a ribosome, and tRNA is formed. This requires several initiation factors (IF1, IF2, IF3, etc., not shown). Then elongation (**2**) follows: a further amino acid, determined by the next codon, is attached. A three-phase elongation cycle develops, with codon recognition, peptide binding of the next amino acid residue, and movement (translocation) of the ribosome three nucleotides further in the 3′ direction of the mRNA. Translation ends with termination (**3**), when one of the three mRNA stop codons (UAA, UGA, or UAG) is reached. Translation (protein synthesis) in eukaryotes occurs outside the cell nucleus in ribosomes in the cytoplasm. The biochemical processes of the stages shown here have been greatly simplified.

D. Structure of transfer RNA

Transfer RNA has a characteristic, cloverleaf-like structure, illustrated here by yeast phenylalanine tRNA (**1**). It has three single-stranded loop regions and four double-stranded "stem" regions. The three-dimensional structure (**2**) is complex, but various functional areas can be differentiated, such as the recognition site (anticodon) for the mRNA codon and the binding site for the respective amino acid (acceptor stem) on the 3′ end (acceptor end).

Further Reading

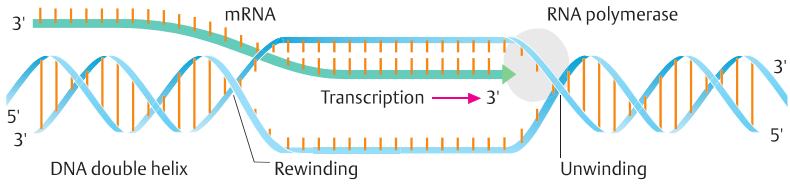
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Brenner S, et al. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 1961; 190: 576–581

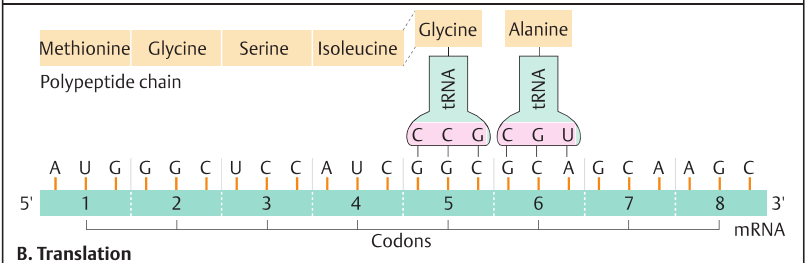
Ibba M, Söll D. Quality control mechanisms during translation. Science 1999; 286(5446): 1893–1897

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

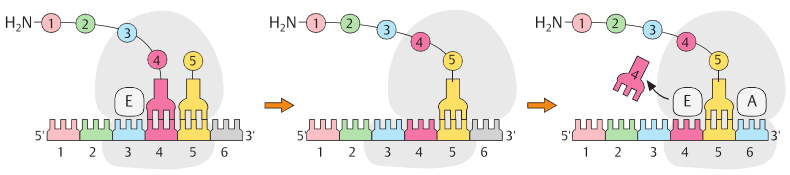
Watson JD, et al. Molecular Biology of the Gene. 6th ed. New York, NY: Cold Spring Harbor Laboratory Press, 2008



A. Transcription

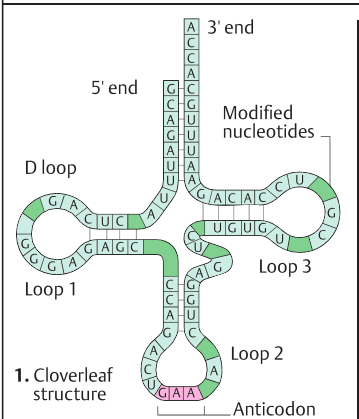


B. Translation

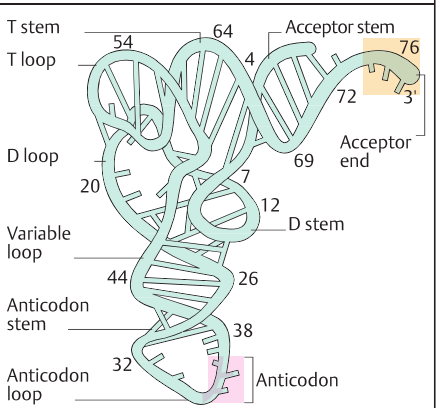


Step 1 | Step 2 Elongation | Step 3

C. Stages of translation



1. Cloverleaf structure



2. Three-dimensional structure

D. Structure of transfer RNA (tRNA)

Genetic Code

The genetic code is a set of biological rules by which DNA nucleotide base-pair sequences are translated into corresponding sequences of amino acids. The genetic code is a triplet code. Each code word (codon) for an amino acid consists of a sequence of three nucleotide base pairs. The genetic code is called *degenerate* because several amino acids are encoded by more than one codon (on average three, but up to six codons for leucine, serine, and arginine). The genetic code also specifies the beginning (start codon) and the end (stop codon) of the coding region. The genetic code is almost universal. Most organisms (bacteria and viruses, animals, and plants) use the same code, except for mitochondria and chloroplasts in plants.

A. Genetic code in messenger RNA

The first, second, and third nucleotide bases in the 5′ to 3′ direction determine the codon for one of the 20 amino acids. Each codon corresponds to one amino acid. The genetic code usually is written in the code language of messenger RNA (mRNA), using uracil (U) instead of thymine (T) in DNA. An example for the redundancy of the genetic code is the amino acid phenylalanine, which is encoded by UUU and UUC. The six codons that define the amino acid serine are UCU, UCC, UCA, UCG, AGU, and AGC. The third position (at the 3′ end of the triplet) of a codon is the most variable. The only amino acids that are encoded by a single codon are methionine (AUG) and tryptophan (UGG). The start codon is AUG (methionine). Stop codons are UAA, UAG, and UGA.

The genetic code was elucidated in 1966 by analyzing how triplets transmit information from the genes to proteins. mRNA added to bacteria could be directly converted into a corresponding protein. Some deviations from the universal genetic code are listed in the table below.

B. Abbreviated code

To save space, long sequences of amino acids are usually written using abbreviations, with each amino acid designated by just one letter.

C. Overlapping reading frames

A reading frame is a nucleotide sequence from the start codon to the stop codon. The interval between a start and a stop codon containing genetic information is called an *open reading frame* (ORF). An ORF does not contain a further stop codon. Normal reading frames do not overlap. Of the possible three reading frames A, B, or C, only one is correct (ORF A). The reading frames B and C are interrupted by a stop codon after three and five codons, respectively, and cannot serve to code sequences.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Crick FH, et al. General nature of the genetic code for proteins. Nature 1961; 192: 1227–1232

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Deviations from the universal genetic code

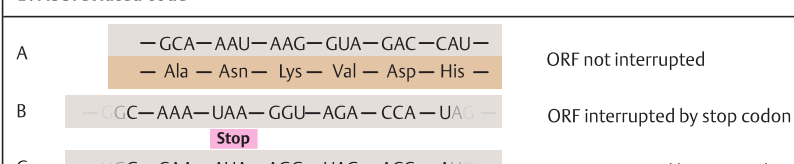
Codon	Universal	Deviation	Occurrence
UGA	Stop	Tryptophan	*Mycoplasma*, mitochondria of some species
CUG	Leucine	Threonine	Mitochondria in yeast
UAA, UAG	Stop	Glycine	*Acetabularia*, *Tetrahymena*, *Paramecium*
UGA	Stop	Cysteine	*Euplotes*

Nucleotide base				
First	Second			Third
	Uracil (U)	Cytosine (C)	Adenine (A)	Guanine (G)	
Uracil (U)	F Phenylalanine (Phe) F Phenylalanine (Phe) L Leucine (Leu) L Leucine (Leu)	S Serine (Ser) S Serine (Ser) S Serine (Ser) S Serine (Ser)	Y Tyrosine (Tyr) Y Tyrosine (Tyr) Stop Codon Stop Codon	C Cysteine (Cys) C Cysteine (Cys) Stop Codon W Tryptophan (Trp)	U C A G
Cytosine (C)	L Leucine (Leu) L Leucine (Leu) L Leucine (Leu) L Leucine (Leu)	P Proline (Pro) P Proline (Pro) P Proline (Pro) P Proline (Pro)	H Histidine (His) H Histidine (His) Q Glutamine (Gln) Q Glutamine (Gln)	R Arginine (Arg) R Arginine (Arg) R Arginine (Arg) R Arginine (Arg)	U C A G
Adenine (A)	I Isoleucine (Ile) I Isoleucine (Ile) I Isoleucine (Ile) Start (Methionine)	T Threonine (Thr) T Threonine (Thr) T Threonine (Thr) T Threonine (Thr)	N Asparagine (Asn) N Asparagine (Asn) K Lysine (Lys) K Lysine (Lys)	S Serine (Ser) S Serine (Ser) R Arginine (Arg) R Arginine (Arg)	U C A G
Guanine (G)	V Valine (Val) V Valine (Val) V Valine (Val) V Valine (Val)	A Alanine (Ala) A Alanine (Ala) A Alanine (Ala) A Alanine (Ala)	D Aspartic acid (Asp) D Aspartic acid (Asp) E Glutamic acid (Glu) E Glutamic acid (Glu)	G Glycine (Gly) G Glycine (Gly) G Glycine (Gly) G Glycine (Gly)	U C A G

A. Genetic code for all amino acids in mRNA

Start AUG	F (Phe) UUU UUC	L (Leu) CUU CUC CUG CUA UUG UUA	R (Arg) CGU CGC CGG CAA AGG AGA	V GUU GUC GUG GUA	
Stop UAA UAG UGA	G (Gly) GGU GGC GGG GGA	M (Met) AUG	S (Ser) UCU UCC UCG UCA AGU AGC	W (Trp) UGG	
A (Ala) GCU GCC GCG GCA	H (His) CAU CAC	N (Asn) AAU AAC		Y (Tyr) UAU UAC	
C (Cys) UGU UGC	I (Ile) AUU AUC AUA	P (Pro) CCU CCC CCG CCA	T (Thr) ACU ACC ACG ACA	B (Asx) Asn or Asp	
D (Asp) GAU GAC				Z (Glx) Gln or Glu	
E (Glu) GAG GAA	K (Lys) AAG AAA	Q (Gln) CAG CAA			

B. Abbreviated code



C. Overlapping reading frames

Eukaryotic Gene Structure

In eukaryotic genes, the coding sequences are interrupted by noncoding sequences of variable length. The coding sequences are called *exons*, and the noncoding sequences are called *introns*, two terms introduced in 1978 by W. Gilbert. The introns are removed before translation can begin. This process is called *RNA processing*. Some introns contain sequences necessary for the regulation of gene activity.

A. Exons and introns

In 1977, it was found unexpectedly that the DNA of a eukaryotic gene is longer than its corresponding messenger RNA (mRNA). When mRNA is hybridized to its complementary single-stranded DNA, loops of single-stranded DNA remain, as shown in the electron micrograph (**1**). mRNA hybridizes only with certain sections of the single-stranded DNA because it is shorter than its corresponding DNA coding strand (**2**). Seven loops (**A–G**) and eight hybridizing sections are shown (**1–7**, and the leading section L). Of the total 7,700 DNA base pairs of this gene (**3**), only 1,825 hybridize with mRNA. Each hybridizing segment is an exon, here a total of seven. The single-stranded segments that do not hybridize correspond to the introns. The size and arrangement of exons and introns are characteristic for every eukaryotic gene (exon/intron structure). (Electron microphotograph Chambon 1981).

B. Intervening DNA sequences (introns)

In prokaryotes, DNA is colinear with mRNA and contains no introns (**1**). In eukaryotes, mature mRNA is complementary to only certain sections of DNA because the latter contains introns (**2**).

C. Basic eukaryotic gene structure and transcript processing

Exons and introns are numbered in the 5′ to 3′ direction of the coding strand. Both exons and introns are transcribed into a precursor RNA (primary transcript). The first and the last exons usually contain sequences that are not translated. These are called the *5′ untranslated region* (5′ UTR) of exon 1, and the 3′ UTR at the 3′ end of the last exon. After transcription, RNA processing begins by removing the noncoding segments (introns) from the primary transcript in a process called *splicing*. Splicing occurs in a spliceosome in a very precise manner in order to maintain the correct reading frame.

Introns almost always start with the nucleotides GT in the 5′ to 3′ strand (GU in RNA) and end with AG. The sequences at the 5′ end of the intron beginning with GT are called the *splice donor site*. The splice acceptor site is at the 3′ end of the intron. Mature mRNA is modified at its 5′ end by adding a stabilizing structure called a *cap*, and at its 3′ end by adding many adenines (polyadenylation).

D. Splicing pathway in GU–AG introns

Splicing is achieved by RNA molecules rather than proteins. A spliceosome consists of five small nuclear ribonucleoproteins (snRNAs) and proteins. Each forms a small nuclear ribonucleoprotein with at least seven protein subunits. Schematically, the basic mechanism of splicing involves autocatalytic cleavage at the 5′ end of the intron, resulting in lariat formation. This is an intermediate circular structure formed by connecting the 5′ terminus (UG) to a base (A) within the intron. This site is called the *branch site*. In the next stage, cleavage at the 3′ site releases the intron in lariat form. At the same time, the right exon is ligated (spliced) to the left exon. The lariat is debranched to yield a linear intron, and this is rapidly degraded. The branch site identifies the 3′ end for precise cleavage at the splice acceptor site and lies 18 to 40 nucleotides upstream (in the 5′ direction) of the 3′ splice site.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

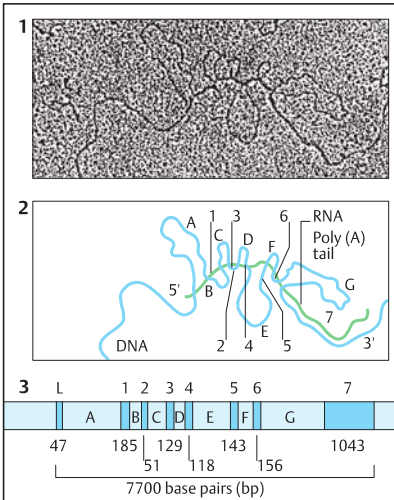
Chambon P. Split genes. Sci Am 1981; 244(5): 60–71

Krebs JE, et al. Kilpatrick: Lewin's Genes XI. Sudbury: Jones & Bartlett, 2014

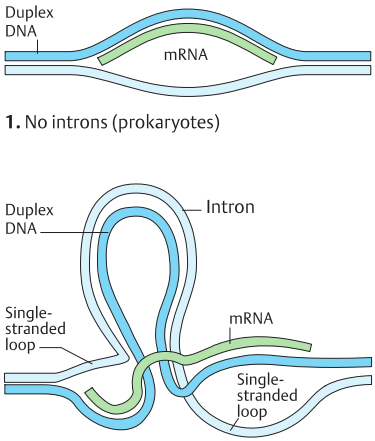
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Watson JD, et al. Molecular Biology of the Gene. 6th ed. New York, NY: Cold Spring Harbor Laboratory Press, 2008



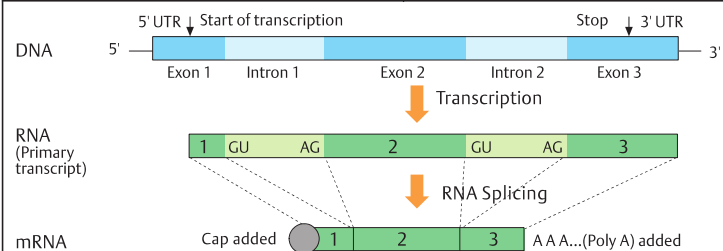
A. Exons and introns



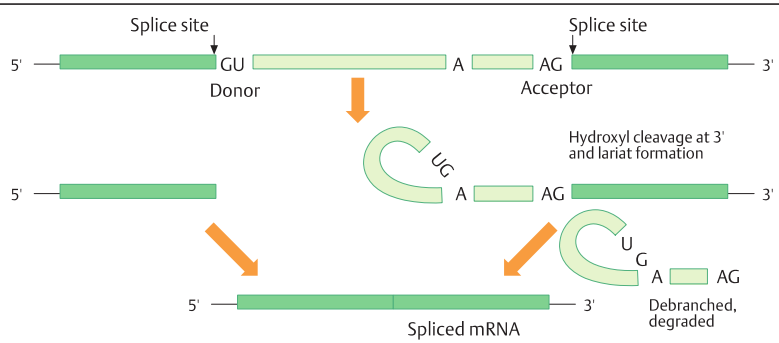
1. No introns (prokaryotes)

2. With introns (eukaryotes)

B. Intervening DNA sequences (introns)



C. Basic eukaryotic gene structure and transcript



D. Splicing pathway in GU-AG introns

Restriction Enzymes

DNA can be cleaved into relatively small fragments in a reproducible manner by enzymes that cleave DNA at specific sites. Each enzyme cleaves only when an enzyme-specific nucleotide sequence occurs (its recognition sequence or restriction site). These enzymes are restriction endonucleases (called *restriction enzymes*) and are present in different bacteria. Approximately 400 different restriction endonucleases have been derived from various bacteria. Restriction endonucleases protect bacteria from invading foreign DNA. The sizes of the DNA fragments produced by a particular enzyme depend on the distribution of the restriction sites.

A. DNA cleavage by restriction nucleases

Many restriction enzymes cleave DNA asymmetrically. This leaves a single-stranded part of the DNA at both ends of a DNA fragment. For example, the recognition site of the common restriction enzyme *Eco*RI, derived from the bacterium *Escherichia coli* restriction enzyme I, is 5′-GAATTC-3′ (**1**). It cleaves asymmetrically producing fragments with overhangs of single-stranded DNA. On one fragment, the single-stranded 3′ to 5′ end has four nucleotides (3′-TTAA) overhanging, and on the other fragment, the 5′ to 3′ overhang is AATT-3′ (**2**). Such an asymmetric cleavage pattern is called *palindromic* because it reads the same in opposite directions. Some restriction enzymes have a symmetric recognition site (**3**) and produce blunted ends (**4**) as the enzyme *Hae*III (5′-CGCG-3′). The ends of fragments with single-stranded overhangs can be easily connected.

B. Examples of restriction enzymes

Restriction enzymes can be classified according to the type of ends they produce: (**1**) 5′ overhangs (e.g., *EcoRI*, discussed earlier), (**2**) 3′ overhangs (e.g., *PstI*), (**3**) blunted ends (e.g., *AluI, HaeIII* [discussed earlier], *HpaI*), or (**4**) nonpalindromic ends (*MlnI*). Some have a bipartite recognition sequence with different numbers of nucleotides at the single-stranded ends (e.g., *BstI*).

In some enzymes as, for example, *HindII*, it suffices that the two middle nucleotides are a pyrimidine and a purine (GTPy-PuAC), and it does not matter whether the former is thymine (T) or cytosine (C) or whether the latter is adenine (A) or guanine (G). Such recognition sites occur frequently and produce many relatively small fragments. Rare cutters recognize long sites of 10 and more nucleotides. Consequently, they produce large fragments, which are useful for many purposes. Some enzymes have cutting sites with limited specificity.

C. Determining of the location of restriction sites

Since the fragment sizes reflect the relative positions of the cleavage sites, they can be used to characterize a DNA segment (restriction map). For example, if a DNA segment of 10 kilobases (kb) is cleaved by two enzymes, A and B, and the result is three fragments of 2, 3, and 5 kb, then the relative locations of the cleavage sites can be determined by using enzyme A and enzyme B. If enzyme A yields two fragments of 3 and 7 kb, and enzyme B yields two fragments of 2 and 8 kb, then the two recognition sites of enzymes A and B must be located 5 kb apart. The interpretation is that the recognition site for A must be 3 kb from the left end and that for B must be 2 kb from the right end (red arrows).

D. Restriction map

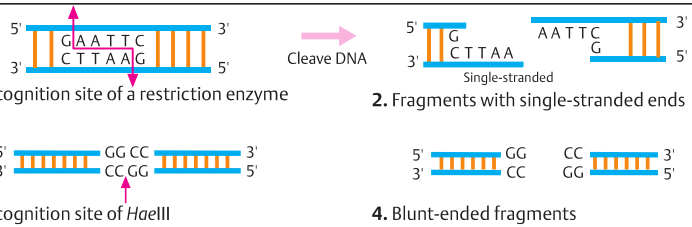
In the example shown, a DNA segment is characterized by the distribution of the recognition sites for enzymes E (*EcoRI*) and H (*HindIII*). The individual sites are separated by intervals defined by the size of the fragments after digestion with the enzyme. Restriction mapping was of considerable importance in medical genetics and evolutionary research prior to the availability of direct information by DNA sequencing.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011



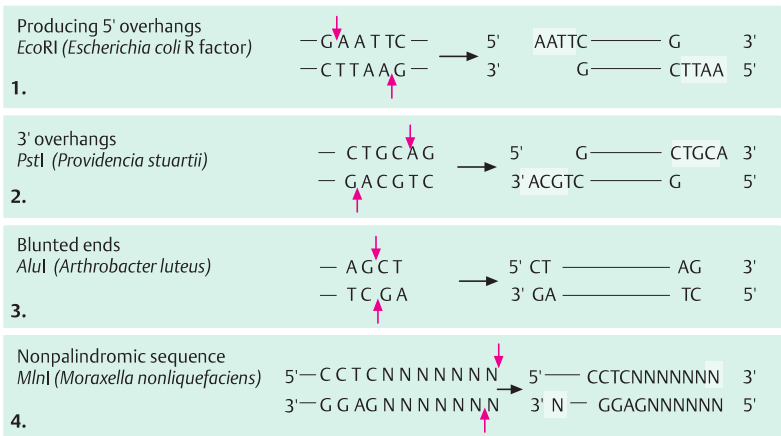
1. Recognition site of a restriction enzyme

2. Fragments with single-stranded ends

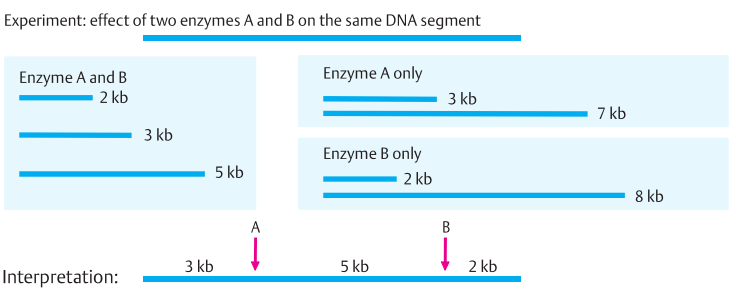
3. Recognition site of *Hae*III

4. Blunt-ended fragments

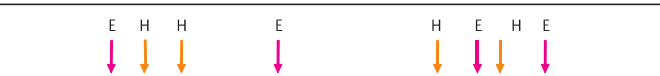
A. DNA cleavage by restriction nucleases



B. Examples of restriction enzymes



C. Determining the location of restriction sites



D. Restriction map

DNA Amplification (PCR)

The introduction, in 1985, of cell-free methods for multiplying DNA fragments of defined origin from a complex mixture has greatly facilitated the molecular analysis of DNA and genes. The polymerase chain reaction (PCR) is a cell-free, rapid, sensitive, automated method for multiplying DNA fragments from minute amounts of DNA. It is so sensitive that contamination with undesired DNA occurs easily, which requires special precautions.

A. Polymerase chain reaction

Standard PCR is an *in vitro* procedure for amplifying defined target DNA sequences from small amounts of DNA. The selective amplification requires prior information about DNA sequences flanking the target DNA. Two oligonucleotide primers of approximately 15 to 25 base pairs in length are used to initiate DNA synthesis. The primers are complementary to sequences outside the 3′ ends of the target site on the two DNA strands and bind specifically to these. During each PCR cycle, double-stranded DNA molecules are denatured, and each single strand is used as a template for synthesis of a new strand and is renatured (annealed) with a complementary strand under controlled conditions. PCR is a chain reaction of approximately 25 to 35 cycles. Each cycle, involving three precisely time-controlled and temperature-controlled reactions performed in an automated thermal cycler, takes approximately 1 to 5 minutes. The three steps in each cycle are as follows: (i) denaturation of double-stranded DNA at approximately 93 to 95°C for human DNA, (ii) primer annealing at approximately 50 to 70°C, depending on the expected melting temperature of the duplex DNA, and (iii) DNA synthesis using heat-stable DNA polymerase (from microorganisms living in hot springs, for example, *Thermophilus aquaticus, Taq polymerase*), typically at approximately 70 to 75°C. At each subsequent cycle, the template (shown in blue) and the DNA newly synthesized during the preceding cycle (shown in red) act as templates for another round of synthesis. The first cycle results in newly synthesized DNA of varied lengths (shown with an arrow) at the 3′ ends because synthesis continues beyond the target sequences. However, DNA strands of fixed length at both ends rapidly outnumber those of variable length because synthesis cannot proceed past the terminus of the primer at the opposite template DNA. At the end, at least 10^5 copies of the specific target sequence are present. This can be visualized as a distinct band of a specific size after using gel electrophoresis. In addition to the standard reaction, a wide variety of automated PCR-based methods have been developed for different purposes. (Figure adapted from Alberts et al, 2015).

B. Reverse transcriptase polymerase chain reaction

The reverse PCR (RT-PCR) approach uses mRNA as starting material. After the first primer is attached, complementary new DNA is synthesized by reverse transcription. This is used as a template for a new DNA strand. Subsequently multiple copies of cDNA are produced by PCR.

C. Allele-specific polymerase chain reaction

This amplifies a DNA sequence from one allele only, and not the other if it is different. If allele 1 contains an A–T base pair at a particular site (**1**) and allele 2 contains a C–G pair (**2**), the two alleles can be distinguished. Allele-specific PCR is normal for allele 1 (**3**). If a mutation has changed in allele 1 the T by a C, amplification is not possible (**4**). Similarly, PCR for allele 2 will be normal (**5**), but not if a C has been replaced by an A in allele 2, allele-specific amplification is not possible (**6**). Reverse transcriptase PCR can be used when the known exon sequences are widely separated within a gene. By using rapid amplification of complementary DNA (cDNA) ends (RACE-PCR), the 5′ and 3′ end sequences can be isolated from cDNA. Other variations of PCR are Alu-PCR, anchored PCR, real-time PCR, and others.

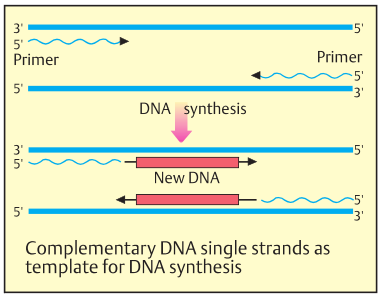
Further Reading

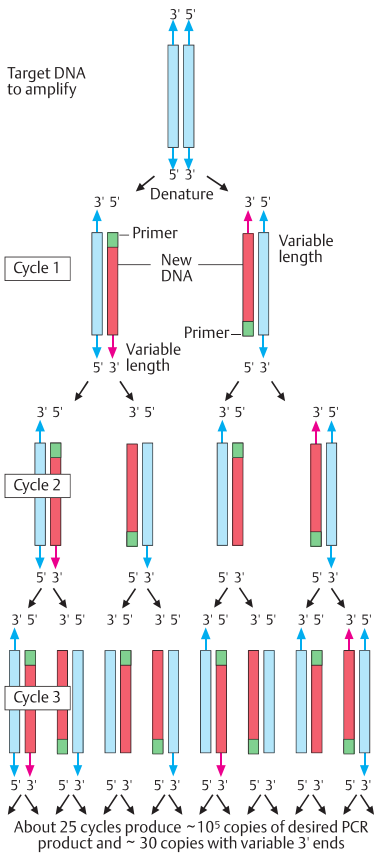
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

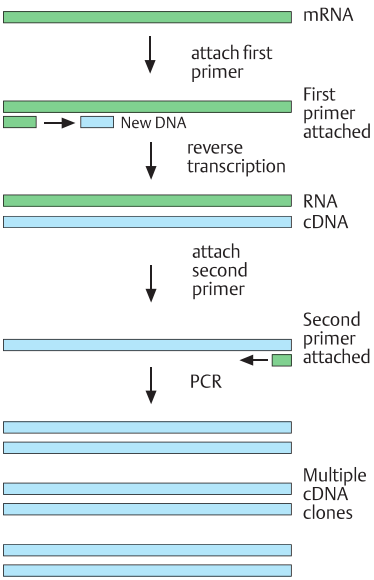
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Strachan T, et al. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

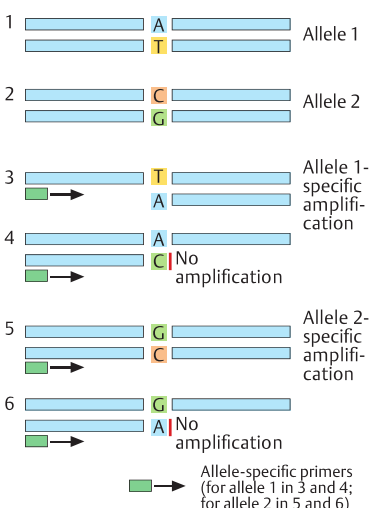




A. Polymerase chain reaction (PCR)



B. Reverse PCR (RT-PCR)



C. Allele -specific PCR

DNA Sequencing

The introduction of two basic methods for determining the sequence of the nucleotide base pairs in (DNA sequencing) in 1977 represented a milestone in genetic analysis. These are based on a chemical cleavage method developed by A. M. Maxam and W. Gilbert, and an enzymatic method developed by F. Sanger and coworkers. The chemical cleavage method is no longer used, whereas the Sanger principle was widely used and eventually applied as the first automated sequencing method.

A. Dideoxy DNA sequencing (Sanger sequencing)

This method rests on the principle that DNA synthesis is terminated when a dideoxynucleotide (ddATP, ddTTP, ddGTP, ddCTP) is used instead of a normal deoxynucleotide (dATP, dTTP, dGTP, dCTP). A dideoxynucleotide (ddNTP) is an analogue of the normal dNTP. It differs by lack of a hydroxyl group at the 3′ carbon position. When a ddNTP is incorporated during DNA synthesis, no bond between its 3′ position and the next nucleotide is possible.

DNA synthesis is initiated using a primer of approximately 20 RNA nucleotides in four parallel reactions containing all four dNTPs and one of the four ddNTPs labeled with ^{32}P in the phosphate groups (**1**). When ddATP is incorporated into the new DNA chain, the further synthesis is inhibited beyond any site containing a T in the template. By setting the concentration of ddATP low enough, the chain will be terminated at all sites after an A is incorporated. The resulting fragments share the same 5′ end but differ in length at their 3′ end, each ending with an A (**1**). Thus, each DNA fragment ends with an A at a defined position. Similarly, in the other three reactions, the growing DNA chain is terminated at a G by ddGTP (**2**), at a C by ddCTP (**3**), and at a T by ddTTP (**4**). By aligning the ends of the fragments according to size, the nucleotides at the ends of the fragments will yield their sequence accurately. Initially, this was done by gel electrophoresis, later by automated sequencing as shown in part B.

B. Automated DNA sequencing

Automated DNA sequencing is based on fluorescence labeling of DNA and suitable detection systems. The direct fluorescent labels are fluorophores, which emit a distinct fluorescent color when exposed to ultraviolet light of a specific wavelength. Each specifically attaches to one of the four nucleotide bases and produces a signal that identifies the base. Thus, each of the four bases can be labeled distinctly. The resulting fluorescent signal is recorded at a fixed point when DNA passes through a capillary containing an electrophoretic gel. Each ddNTP is labeled with a different color, for example, ddATP green, ddCTP blue, ddGTP black, and ddTTP red (**1**). (The actual colors for each nucleotide may be different.) All chains terminated at an adenine (A) will yield a green signal, as in this example, all chains terminated at a cytosine (C) will yield a blue signal, and so on. The sequencing reactions (**2**) are performed in sequencing capillaries (**3**). The electrophoretic migration of the ddNTP-labeled chains in the gel in the capillary passes in front of a laser beam focused on a fixed position (**4**). The sequence is electronically read and recorded in a detection system, and is visualized as peaks in one of the four colors, representing the alternating nucleotides in their sequence positions (**5**). Several variant techniques for sequencing have been devised. (Print of sequence provided by N. Wagner, Essen).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Pettersson E, et al. Generations of sequencing technologies. Genomics 2009; 93(2): 105–111

Sanger F, et al. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977; 74 (12): 5463–5467

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

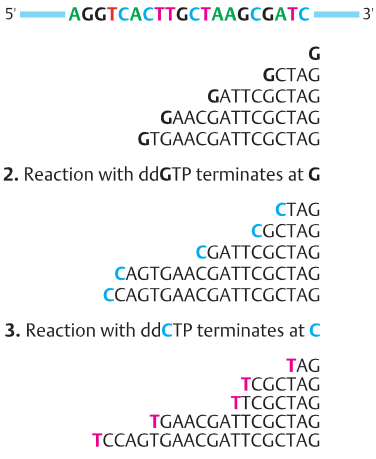
Single-stranded template DNA

1. Reaction with ddATP terminates synthesis at **A**

Sequence obtained from all four reactions

3' TCCAGTGAACGATTCGCTAG 5'

(reverse of template)



2. Reaction with dd**G**TP terminates at **G**

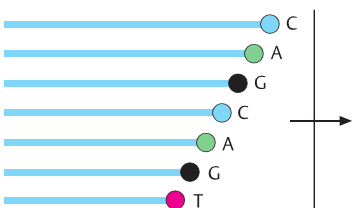
3. Reaction with dd**C**TP terminates at **C**

4. Reaction with dd**T**TP terminates at **T**

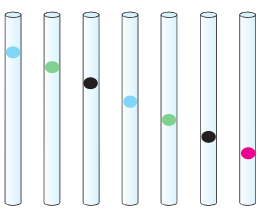
A. Dideoxy DNA sequencing

ddATP ddCTP ddGTP ddTTP

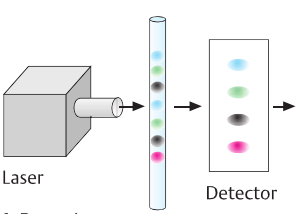
1. ddNTPs each with different fluorescent dye



2. Sequencing reactions

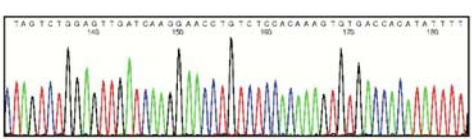


3. Capillaries containing sequencing gel with NTPs



4. Detection system

Detector



5. Printout of sequence obtained by Sequence Analyser. The numbers indicate nucleotide positions. (T red, A green, G black, C blue)

B. Automated DNA sequencing

Parallel DNA Sequencing (Next-Generation Sequencing)

Since 2005, several completely new DNA sequencing technologies have been introduced. They are also referred to as next-generation sequencing. These approaches differ fundamentally from dideoxy sequencing according to the Sanger principle (previous plate). They consist of a combination of template preparation, sequencing by DNA synthesis without electrophoresis, high-resolution imaging, sequence tag assembly, and genome alignment methods. Millions of sequences can be automatically determined in parallel to yield whole-genome sequences in a relatively short time and at reduced costs (whole-genome sequencing). The methods differ and yield different read length and sequenced output. Approximately 1 to 50 gigabases (10^9) can be generated in a matter of days. The principle of one selected example is presented here.

A. Selection of ligated DNA fragments

Double-stranded DNA is first fragmented (**1**), and after end repair, overhangs of adenine (A) molecules are added (**2**). A unique adapter sequence is ligated to each fragment (**3**). This creates a library with a selection of ligated DNA (**4**).

B. Cluster generation and polony array formation

Adapter-modified single-stranded DNA fragments are attached to a fluid channel surface within a glass slide, called a *flow cell*. Solid-phase bridge amplification is achieved with an immobilized template of dNTPs, and DNA polymerase. This results in multiple (100–200 million) spatially separated template clusters. Following denaturation into single-stranded DNA, each member of this population of templates will undergo an individual sequencing reaction. This starts at a free end, to which a universal sequencing primer can be annealed to initiate the sequencing reaction. Attached to an array, polymerase-driven colonies of growing DNA strands (termed *polonies*) result.

C. Sequencing by synthesis and imaging

Sequencing is achieved by continuous DNA synthesis rather than dideoxy chain termination as in the Sanger principle. Each synthesis reaction is visualized by four colors, each representing one of the four nucleotides synthesized. From the template of the sequence CGAGGC, the synthesized sequence is generated (GCTCC shown).

D. Overview

Four principle phases are DNA fragmentation (**1**), in vitro adapter ligation (**2**), generation of polony array (**3**), and cycle array sequencing (**4**). (Figures derived from a poster that was kindly provided by Illumina Inc., San Diego, California, United States.)

Exome sequencing: this automated approach is directed at all coding sections of the genome, the exons. Starting with genomic DNA fragments, a library is established. The library fragments are flanked by adapter oligonucleotides, and exome-containing fragments are captured and enriched. The captured DNA fragments derived from the exons are aligned and sequenced.

Further Reading

Bamshad MJ, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011; 12(11): 745–755

Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008; 456(7218): 53–59

Mardis ER. A decade's perspective on DNA sequencing technology. Nature 2011; 470(7333): 198–203

Margulies M, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005; 437(7057): 376–380

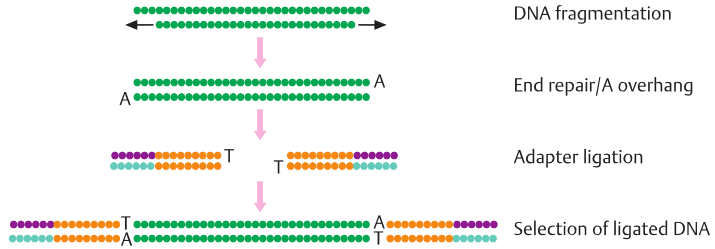
Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010; 11(1): 31–46

Pettersson E, et al. Generations of sequencing technologies. Genomics 2009; 93(2): 105–111

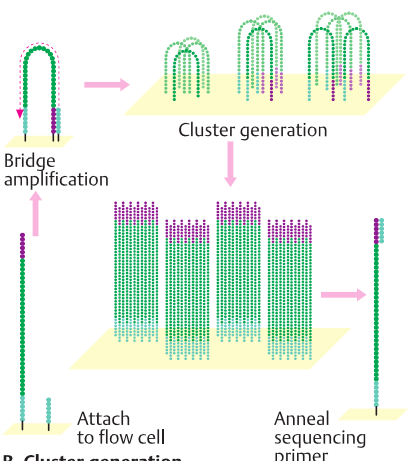
Shendure J, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 2005; 309(5741): 1728–1732

Strachan T, et al. Genetisc and Genomics in Medicine. New York, NY: Garland Science, 2015

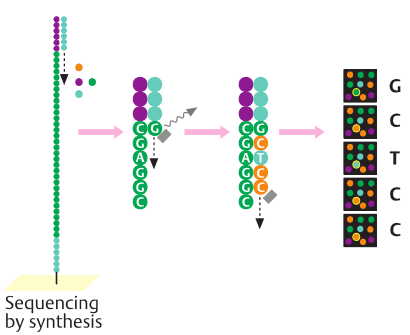
Voelkerding KV, et al. Next-generation sequencing: from basic research to diagnostics. Clin Chem 2009; 55(4): 641–658



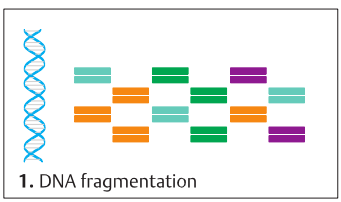
A. Selection of ligated DNA fragments

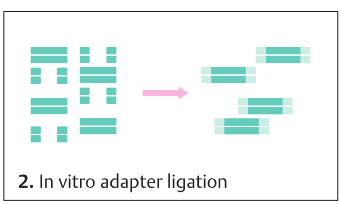


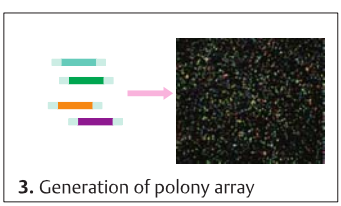
B. Cluster generation

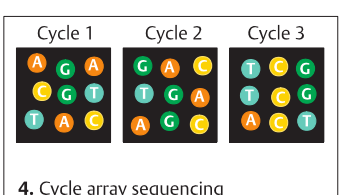


C. Sequencing by synthesis and imaging









D. Overview

DNA Cloning

DNA cloning selectively amplifies specific DNA sequences as fragments. The correct DNA fragments are identified by hybridization of complementary single-stranded DNA. A short segment of single-stranded DNA (a DNA probe) originating from the sequence to be studied will hybridize to its complementary single-stranded sequences. This serves to identify and select the fragment under study. After the hybridized sequence has been separated from other DNA, it can be cloned.

A. Principle of cell-based DNA cloning

Cell-based DNA cloning requires a series of consecutive steps. First, a collection of different DNA fragments (labeled 1, 2, and 3) are obtained from the desired DNA (target DNA) by cleaving it with a restriction enzyme (see p. 66) (**1**). Fragments with short single-stranded ends resulting from restriction enzyme cleavage are ligated to DNA fragments containing the origin of replication (OR) of a replicon. This enables them to replicate in the presence of a selectable marker, for example, a DNA sequence containing an antibiotic resistance gene (**2**). The recombinant DNA molecules are transferred into host cells (bacterial or yeast cells), where they can replicate independently of the host cell genome (**3**). Host cells that have incorporated the fragment to be cloned will be transformed (here fragment 1 in a brown circle). Usually, the host cell takes up only one foreign DNA molecule. The host cells transformed by recombinant (foreign) DNA are grown in culture, where they multiply (propagation, **4**). Selective growth of transformed cells containing the desired DNA fragment produces multiple copies of the fragment (**5**). The resulting recombinant DNA clones form a homogeneous population (**6**). They are used to build a large collection of cloned DNA fragments, called a *clone library* (**7**) (see the next page). In cell-based cloning, the replicon-containing DNA molecules are referred to as vector molecules. (Figure adapted from Strachan & Read, 2011).

B. Vector molecules for DNA cloning

Cloning vectors are molecules that allow selection for the DNA fragments sought.

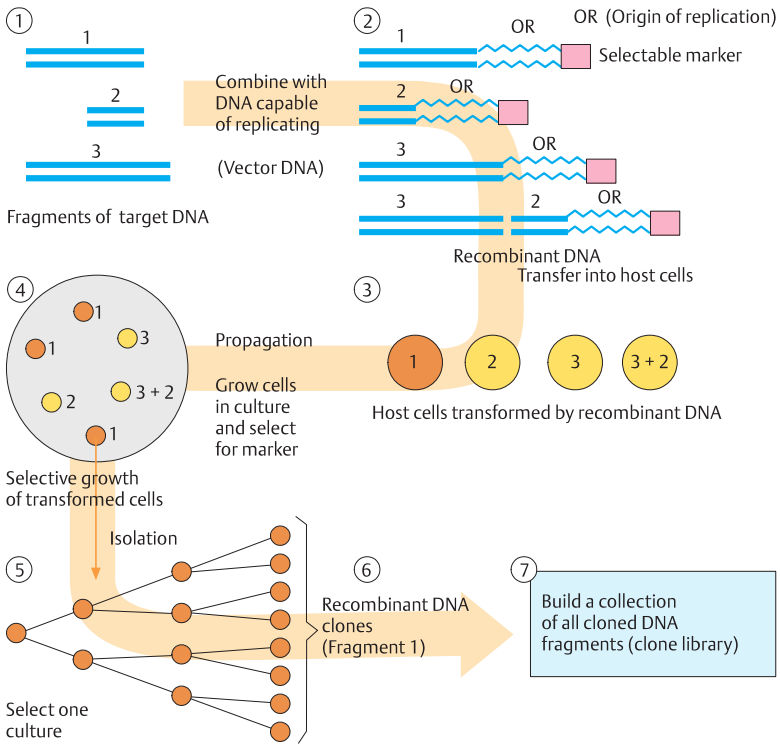
Many different DNA vector systems exist for cloning DNA fragments of different sizes. Plasmid vectors are used to clone small fragments. The experiment is designed so that the plasmid incorporating the fragment to be cloned confers antibiotic resistance to its bacterial host when grown in culture medium containing the antibiotic. One of the first vectors to be developed was the plasmid vector pBR322 (Bolivar et al., 1977). It is small, 4,363 base pairs, and has an OR and two genes encoding antibiotic resistance against ampicillin and tetracycline (**1**). It was constructed to contain recognition sites for several restriction enzymes, as shown. Resistance to two antibiotics (here ampicillin and tetracycline) provides selectable markers. They distinguish cells containing recombinant from those containing nonrecombinant pBR322 plasmids. Bacteria that have been exposed to pBR322 are plated on a medium containing both antibiotics. Only bacteria containing plasmids that have taken up new DNA at the *BamHI* recognition site lose the tetracycline resistance because it is inserted at the site for this resistance (insertional activation) (**2**). Cells containing the nonrecombinant pBR322 plasmid remain resistant to tetracycline. If the enzyme *PstI* is used to incorporate a fragment, ampicillin resistance is lost (the bacterium becomes ampicillin sensitive), but tetracycline resistance is retained. Thus, with the help of replica plating, recombinant plasmids containing the DNA fragment to be cloned can be distinguished from nonrecombinant plasmids by altered antibiotic resistance. Cloning in plasmids (bacteria) has become less important since much larger yeast artificial chromosomes have become available for cloning relatively large DNA fragments.

Further Reading

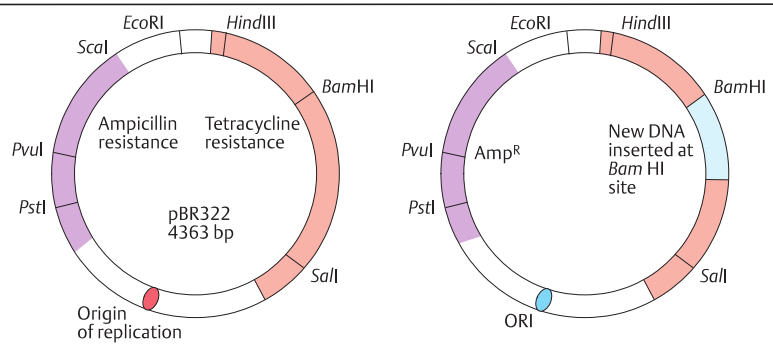
Bolivar F, et al. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 1977; 2(2): 95–113

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011



A. Principle of cell-based DNA cloning



1. Resistant to ampicillin and tetracycline

2. Resistant to ampicillin, sensitive to tetracycline

B. A vector for DNA cloning

DNA Libraries

A DNA library is a random collection of individual DNA fragments that in their entirety represent part or all of the genome of an organism. It was an indispensible approach in the early days of recombinant DNA technologies. Two types of DNA libraries exist: (i) genomic DNA libraries and (ii) complementary DNA (cDNA) libraries. The first is a collection of DNA clones containing all unique nucleotide sequences of the genome. A sufficient number of clones must be present so that every segment of the genome is represented at least once. A cDNA library is a collection of cDNA that is representative of all messenger RNA (mRNA) molecules produced by the organism.

A. Genomic DNA library

Clones of genomic DNA are copies of DNA fragments from all of the chromosomes (**1**). They contain coding and noncoding sequences. Restriction enzymes cleave the genomic DNA into many fragments. Four fragments containing two genes, A and B, are shown schematically (**2**). Fragments of the genes and their surrounding DNA are incorporated into vector molecules, for example, into phage DNA, and are cloned in bacteria or yeast cells. In eukaryotes, a genomic library will contain hundreds of thousands to more than a million of individual DNA clones. A screening procedure is required to find a particular gene (see **C**).

B. cDNA library

A cDNA library consists only of coding DNA sequences. Therefore, it is smaller than a genomic library. The starting material is usually total RNA from a specific tissue or a particular developmental stage of embryogenesis. A limitation is that mRNA can be obtained only from cells in which the respective gene is transcribed, that is, in which mRNA is produced (**1**). Unlike a genomic library, which is complete and contains coding and noncoding DNA, a cDNA library contains only coding DNA. This specificity offers considerable advantages over genomic DNA. However, it requires that mRNA is available and that it does not yield information about the structure of the gene. In eukaryotes, the primary transcript RNA undergoes splicing to form mRNA (**2**; see p. 56). cDNA is derived from mRNA by reverse transcription (**3**). The subsequent steps, that is, incorporation into a vector and replication in bacteria, correspond to those of the procedure for producing a genomic library. The sequence of amino acids in a protein can be determined from cloned and sequenced cDNA. Furthermore, large amounts of a protein can be produced when the cloned gene is expressed in bacteria or yeast cells (*proteome library*).

C. Screening of a DNA library

To identify clones carrying a gene, parts of a gene, or another DNA region of interest, a screening procedure is required. Two principal approaches can be used: (i) detection using oligonucleotide probes that bind to the sequences of interest and (ii) detection based on protein produced by the gene in question. Screening with oligonucleotide probes depends on the principle of hybridization. Single-stranded DNA or RNA molecules will specifically hybridize to their complementary single-stranded sequences (see p. 48). In the membrane-based assay, colonies of cultured bacteria (**1**), some of which will have taken up a recombinant vector, are transferred to a filter paper or a membrane (**2**), lysed, and their DNA is denatured (made single stranded). Labeled complementary single-stranded DNAs or RNAs are used as a probe. The probe will hybridize with only cDNA or RNA (4). After hybridization, a signal appears on the membrane (**5**). DNA complementary to the labeled probe is located here; its exact position in the culture corresponds to that of the signal on the membrane (**5**). A sample is taken from the corresponding area of the culture (**6**). Bacteria from such colonies that have taken up the vectors are grown on agar in a Petri dish to produce multiple copies (clone) of the desired DNA fragments.

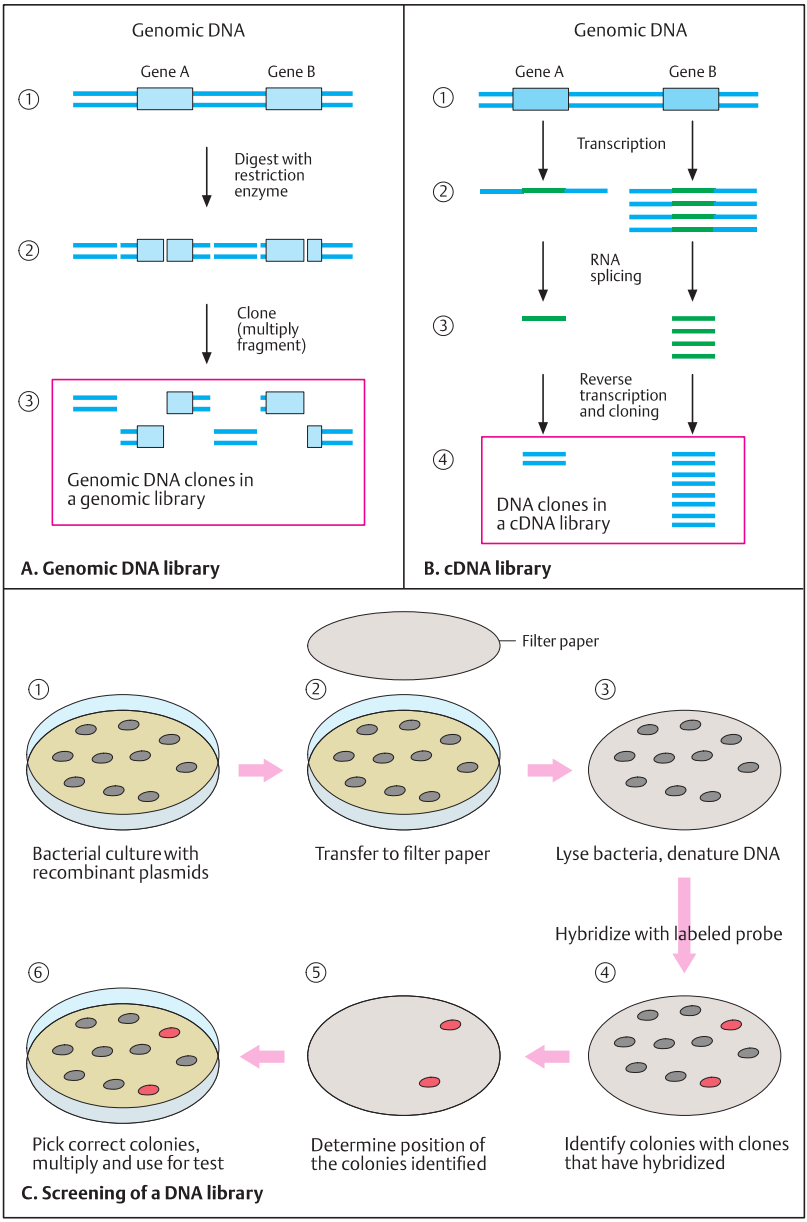
Further Reading

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015

Watson JD, et al. Recombinant DNA. 2nd ed. New York, NY: Scientific American Books, 1992


Genomic DNA
Gene A
Gene B
①
Digest with restriction enzyme
②
Clone (multiply fragment)
③
Genomic DNA clones in a genomic library
A. Genomic DNA library
Genomic DNA
Gene A
Gene B
①
Transcription
②
RNA splicing
③
Reverse transcription and cloning
④
DNA clones in a cDNA library
B. cDNA library
Filter paper
①
Bacterial culture with recombinant plasmids
②
Transfer to filter paper
③
Lyse bacteria, denature DNA
Hybridize with labeled probe
④
Identify colonies with clones that have hybridized
⑤
Determine position of the colonies identified
⑥
Pick correct colonies, multiply and use for test
C. Screening of a DNA library

Southern Blot Hybridization

Southern blot refers to a sensitive method of detecting one or more specific DNA fragments within a complex, random mixture of other DNA fragments. The procedure is named after E. M. Southern, who developed this method in 1975. A corresponding method for assaying RNA is called *Northern blot hybridization* (a word play on Southern). Immunoblotting (Western blot) detects proteins by an antibody-based procedure. Since each restriction enzyme cleaves DNA only at its specific recognition sequence, its recognition sites are distributed unevenly and the DNA fragments differ in size. This method today is rarely employed.

A. Principle of Southern blot hybridization

Total DNA is extracted from white blood cells or other cells (**1**). The DNA is isolated and digested with a restriction enzyme (**2**). The fragments are sorted by size in a gel (usually agarose) in an electric field by using electrophoresis (**3**). The smallest fragments migrate fastest from the cathode to the anode, and the largest fragments migrate slowest. Next, the blot is performed: the fragments contained in the gel are transferred to a nitrocellulose or nylon membrane (**4**). The DNA is denatured (made single-stranded) with alkali and fixed to the membrane by moderate heating (80°C) or ultraviolet cross-linkage. The sample is incubated with a probe of single-stranded DNA (genomic DNA or cDNA) complementary to the region of the gene to be studied (**5**). The probe hybridizes solely with the complementary fragment being sought, not with others. When the probe is radiolabeled, the fragment being sought induces a signal on an X-ray film placed on the membrane. Here it becomes visible as a black band on the film after development (autoradiogram) (**6**).

B. Restriction fragment length polymorphism

Differences in individual DNA sequences result in differences in the location of the recognition sequences of restriction enzymes. This is apparent from differences in sizes of DNA fragments obtained by Southern hybridization (restriction fragment length polymorphism, RFLP). RFLP results when an individual difference of this type creates or eliminates a restriction enzyme recognition site. If an additional site is created, two smaller-than-usual fragments appear in the Southern blot hybridization. If a site is eliminated, one larger-than-usual fragment appears instead of two smaller ones.

An example is shown for a 5-kilobase (kb; 5,000 base pairs) stretch of DNA. On the left (allele 1) it contains a polymorphic restriction recognition site in the middle that is not present in allele 2 on the right. The first allele, with the polymorphic site, is arbitrarily called *allele 1*; the one at the right without this additional site is called *allele 2*.

Southern blot analysis distinguishes the two alleles. On the left, the restriction enzyme cleaves the 5-kb stretch into two fragments. A probe bridging the site will hybridize to both fragments: one 3 kb and the other 2 kb. On the left, a single 5-kb fragment will result. Thus, three possibilities, called the *genotypes*, can be distinguished: (i) two copies of allele 1, homozygous 1–1; (ii) one allele 1 and one allele 2, heterozygous; and (iii) two alleles 2, homozygous 2–2 (for explanations of the terms *homozygous* and *heterozygous*, see pp. 106–114). This approach can be used for indirect detection of a disease-causing mutation. The RFLP itself is unrelated to the mutation. It simply distinguishes DNA fragments of different sizes from the same region. These can be used as markers within a family to determine who is likely to carry a disease-causing mutation and who is not.

Medical relevance

Southern blot hybridization is employed to detect alleles with disease-causing expanded nucleotide repeats (see p. 84, and pp. 254–257).

Further Reading

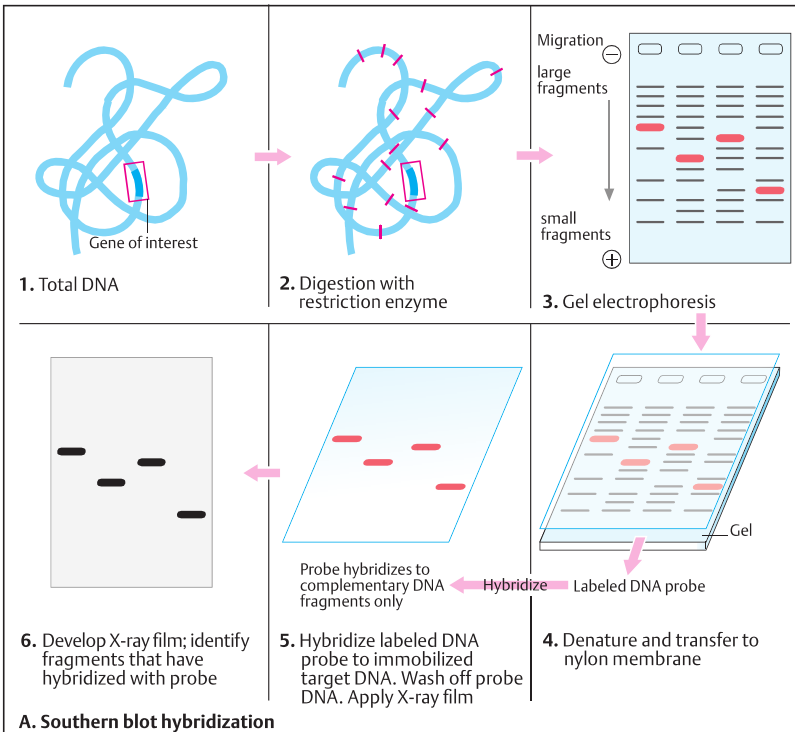
Botstein D, et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 1980; 32(3): 314–331

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

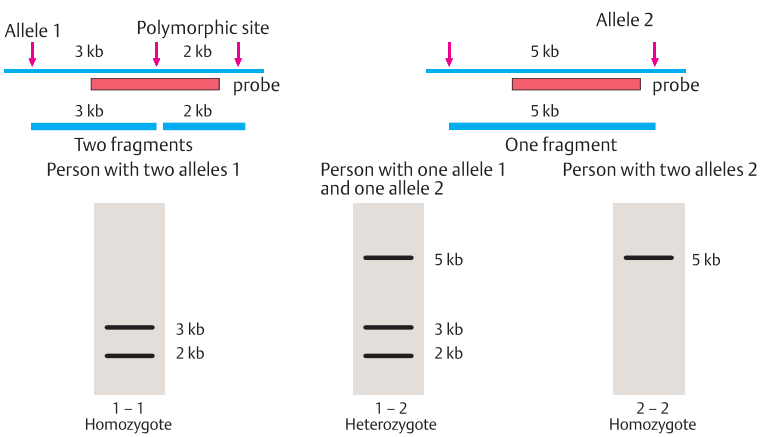
Kan YW, Dozy AM. Antenatal diagnosis of sickle-cell anaemia by D.N.A. analysis of amniotic-fluid cells. Lancet 1978; 2(8096): 910–912

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015



A. Southern blot hybridization



B. Restriction fragment length polymorphism (RFLP)

DNA Variants

Each individual genome differs from any other by differences in the DNA nucleotide sequence. If the variant has a population frequency of 0.01 (1%) or more, it is called a *DNA polymorphism*. On average, approximately 1 in 300 nucleotides is polymorphic. Three basic types of DNA polymorphisms occur: (i) a difference in a single nucleotide (single nucleotide polymorphism, SNP), (ii) small scale variants involving the presence or absence of repeated nucleotides (microsatellites), and (iii) tandem repeats of 10 to 50 nucleotides (minisatellites). The term *satellite* is derived from extra bands present in addition to the bulk of DNA when subjected to density gradient sedimentation in an ultracentrifuge.

A. Single nucleotide polymorphism

Each SNP is identified by a unique identifier, such as rs312476 (where rs stands for reference SNP, followed by a six-digit unique serial number). At any particular site, an SNP involving, for example, adenine (A) and guanine (G) can occur in one of three combinations, AA, AG, or GG, inherited from paternal or maternal chromosome (see Formal Genetics). SNPs can be detected by methods based on a polymerase chain reaction (PCR; see p. 56), which does not require gel electrophoresis.

B. SNP, microsatellite, minisatellite

The three types of common DNA polymorphisms are shown: SNP, microsatellites, and minisatellites. Microsatellites are variable blocks of short tandem repeats of nucleotide sequences. For example, CA repeats can occur with three repeats (5′-CACACA-3′), five repeats, (5′-CACACACACA-3′), six repeats, and so on. Each repeat defines an allele by its number of repeats, for example, three and four.
Minisatellites (also called *variable number of tandem repeats*) consist of repeat units of 20 to 500 base pairs (bp). The size differences resulting from the number of repeats are determined by PCR. These allelic variants differing in the number of tandemly repeated short nucleotide sequences usually occur in noncoding DNA. They are referred to as repetitive DNA (see Part II, Genomics).

C. Genetic variability along a stretch of 100,000 bp

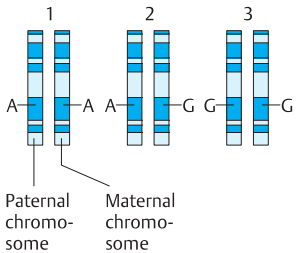
Along a typical stretch of DNA, most variability is represented by SNPs. Minisatellites are quite unevenly distributed and vary in density. (Figure adapted from Cichon et al, 2002).

D. CEPH family

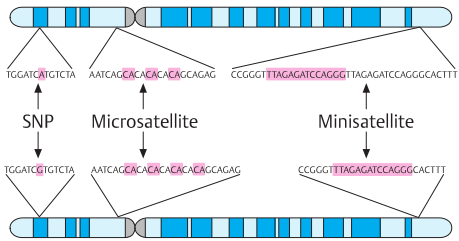
The inheritance patterns of DNA polymorphisms are best recognized in a collection of three-generation families with at least eight children in the third generation. DNA from such families has been collected by the Centre d'Étude du Polymorphisme Humain (CEPH) in Paris, now called the *Centre Jean Dausset*, after the founder. Immortalized cell lines are stored from each family. A CEPH family consists of four grandparents, two parents, and eight children. The schematic figure shows the restriction fragment length polymorphism patterns of a family with four grandparents, two parents, and eight offspring. The four alleles present at a given locus analyzed by Southern blot are designated A, B, C, and D. Starting with the grandparents, the inheritance of each allele through the parents to the grandchildren can be traced. Of the four grandparents, three are heterozygous (AB, CD, BC) and one is homozygous (CC). Since the parents are heterozygous for different alleles (AD father, and BC mother), all eight children are heterozygous: BD, AB, AC, or CD.

Further Reading

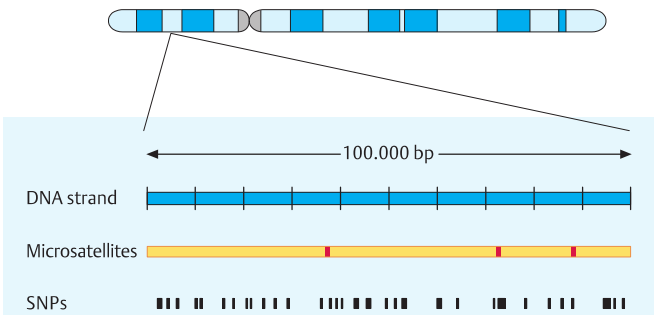
Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007
Cichon S, et al. Variabilität im menschlichen Genom. Dtsch Arztebl 2002; 99:A3091–A3101
Dausset J, et al. Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. Genomics 1990; 6(3): 575–577
Feuk L, et al. Structural variation in the human genome. Nat Rev Genet 2006; 7(2): 85–97 (with online links to databases)
Hinds DA, et al. Whole-genome patterns of common DNA variation in three human populations. Science 2005; 307(5712): 1072–1079
Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013
Strachan T, Read A. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2011



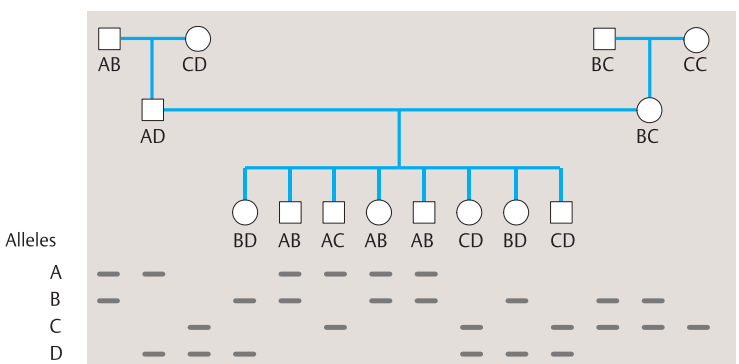
A. Single nucleotide polymorphism



B. SNP, microsatellite, minisatellite



C. Genetic variability along a stretch of 100 000 bp



Alleles A, B, C, D at a marker locus show all possible pairwise combinations

D. CEPH family

Genes and Mutation

Mutation, a term introduced by H. de Vries in 1901, refers to processes that alter the structure and biological function of a gene. The discovery that mutations also occur in bacteria and other microorganisms paved the way to understanding how genes and mutations are related.

When it was recognized that changes (mutations) in genes occur not only spontaneously (T. H. Morgan in 1910) but can also be induced by X-rays (H. J. Muller in 1927), the mutation theory of heredity became a cornerstone of early genetics. The study of mutations is important for several reasons. Mutations cause diseases, including all forms of cancer. Without mutations, well-organized forms of life would not have evolved.

A. Transcription and translation in prokaryotes and eukaryotes

The effects of mutations differ in unicellular and multicellular organisms. In cells without a nucleus, such as bacteria (prokaryotes, **1**), and multicellular organisms (eukaryotes, **2**) with a cell nucleus, their transcription pattern differs. In prokaryotes, the messenger RNA (mRNA) serves directly as a template for translation. The sequences of DNA and mRNA correspond in a strict 1:1 relationship, that is, they are colinear. In eukaryotic cells, a primary transcript of RNA is formed first. The mature mRNA is formed by removing noncoding sections from the primary transcript before it leaves the nucleus to act as a template for the synthesis of a polypeptide (see RNA processing, p. 56).

B. Mutations have a defined site

The systematic analysis of mutations in microorganisms provided the first evidence that coding DNA and its corresponding polypeptides are colinear. Yanofsky et al. (1964) showed that the position of the mutation in the *Escherichia coli* gene encoding the protein tryptophan synthetase A corresponds to the position of the resulting change in the sequence of amino acids. Mutations are shown at four positions. At position 22, phenylalanine (Phe) is replaced by leucine (Leu); at position 49, glutamic acid (Glu) is replaced by glutamine (Gln); and at position 177, Leu is replaced by arginine (Arg). Each mutation has a defined position. Whether it leads to incorporation of a different amino acid depends on how the corresponding codon has been altered. Different mutations at one position (one codon) in different DNA molecules are possible: two different mutations were observed at position 211: glycine (Gly) to arginine (Arg), and Gly to glutamic acid (Glu). Normally (in the wild-type), codon 211 is GGA and codes for glycine. A mutation of GGA to AGA leads to a codon for arginine; a mutation to GAA leads to a codon for glutamic acid.

C. Basic types of mutations

Three different types of mutation, that is, a change from the usual or so-called wild-type, involving single nucleotides (point mutations) can be distinguished: (i) *substitution* (exchange of one nucleotide base for another, altering a codon), (ii) *deletion* (loss of one or more bases), and (iii) *insertion* (addition of one or more bases). Two types of substitution are distinguished: *transition* (exchange of one purine for another purine or of one pyrimidine for another pyrimidine) and *transversion* (exchange of a purine for a pyrimidine or vice versa). A substitution may alter a codon so that a wrong amino acid is present at this site, but without changing the reading frame (*missense mutation*). A deletion or insertion causes a shift of the reading frame (*frameshift mutation*). Thus, the sequence that follows no longer corresponds to the normal sequence of codons. No functional gene product is produced (*nonsense mutation*).

D. Different mutations at the same site

Different mutations may occur at the same site. In the example in **B** (position 211), glycine is replaced by either arginine or glutamic acid.

Further Reading

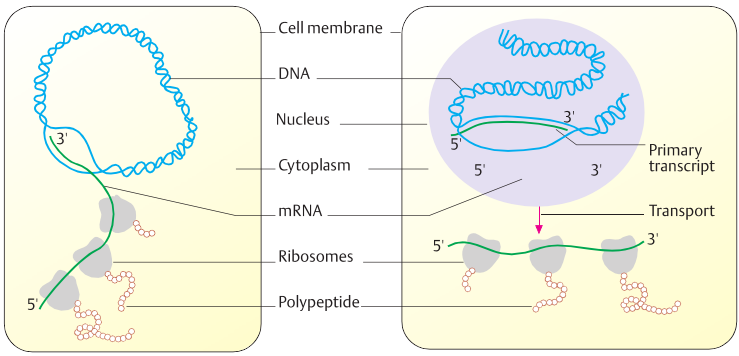
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

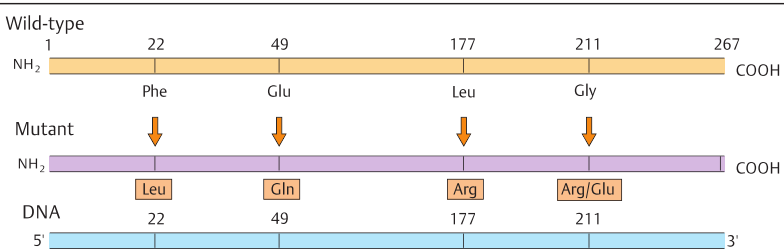
Shendure J, et al. The origins, determinants, and consequences of human mutations. Science 2015; 349 (6255): 1478–1483

Watson JD, et al. Molecular Biology of the Gene. 6th ed. New York, NY: Cold Spring Harbor Laboratory Press, 2008

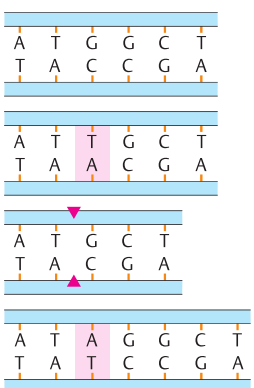
Yanofsky C, et al. On the colinearity of gene structure and protein structure. Proc Natl Acad Sci U S A 1964; 51: 266–272



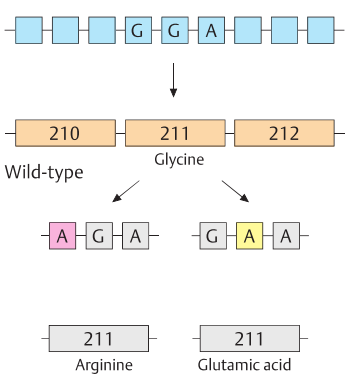
A. Transcription and translation in prokaryotes and eukaryotes



B. Mutations have a defined site



C. Basic types of mutations



D. Different mutations at the same site

Mutations Due to Base Modifications

A mutation can result from a chemical or a physical event that leads to modification of a nucleotide base. If this affects the base-pairing pattern, it will interfere with replication or transcription. A chemical substance able to induce such changes is called a *mutagen*. Mutagens cause mutations in different ways. Spontaneous oxidation, hydrolysis, uncontrolled methylation, alkylation, and ultraviolet (UV) irradiation result in alterations that modify nucleotide bases. DNA-reactive chemicals change nucleotide bases into different chemical structures or remove a base.

A. Deamination and methylation

Cytosine, adenine, and guanine each contain an amino group. When this group is removed (deamination), a modified base with a different base-pairing pattern is the result. Nitrous acid typically removes the amino group. Spontaneous deamination of cytosine occurs at a rate of 100 bases per cell per day (Alberts et al., 2015). This removes the amino group at position 4 (**1**). The resulting molecule is uracil (**2**), and this pairs with adenine rather than guanine. Normally, this is efficiently repaired by uracil-DNA glycosylase. Methylation of the carbon atom at position 5 of cytosine results in 5-methylcytosine, which contains a methyl group at position 5 (**3**). Deamination of 5-methylcytosine will result in a change to thymine, containing an oxygen molecule at position 4 instead of an amino group (**4**). Methylation occurs at a rate of approximately 100 bases per cell per day (Alberts et al., 2015). This mutation will not be corrected because thymine is a natural base. Adenine (**5**) can be deaminated at position 6 to form hypoxanthine, which contains an oxygen molecule in this position instead of an amino group (**6**) and which pairs with cytosine instead of thymine. The resulting change after DNA replication is a cytosine instead of a thymine in the mutant strand.

B. Depurination

Approximately 18,000 purine bases (adenine and guanine) are lost per day from DNA in each cell (depurination) owing to thermal fluctuations (Alberts et al., 2015). Depurination of DNA involves hydrolytic cleavage of the *N*-glycosyl linkage of deoxyribose to the guanine nitrogen in position 9. This leaves a depurinated sugar. The loss of a base pair will lead to a deletion after the next replication if not repaired in time (see DNA repair, p. 74).

C. Alkylation of guanine

Alkylation is the introduction of a methyl or an ethyl group into a molecule. The alkylation of guanine involves reaction with the ketone group at position 6 to form 6-methylguanine. This cannot form a hydrogen bond and thus is unable to pair with cytosine. Instead, it will pair with thymine. Thus, after the next replication, the opposite cytosine (C) is replaced by a thymine (T) in the mutant daughter molecule. As a result, this molecule contains an abnormal GT pair instead of GC. Important alkylating agents are ethylnitrosourea, ethylmethane sulfonate, dimethylnitrosamine, and *N*-methyl-*N*-nitro-*N*-nitrosoguanidine.

D. Base analogues

Base analogues are purines or pyrimidines that are similar enough to the regular DNA nucleotide bases to be incorporated into the new strand during replication. 5-Bromodeoxyuridine is an analogue of thymine. It contains a bromine atom instead of the methyl group in position 5. Thus, it can be incorporated into the new DNA strand during replication. However, the presence of the bromine atom causes ambiguous and often incorrect base pairing.

E. Ultraviolet-light-induced thymine dimers

UV irradiation at 260-nm wavelength induces covalent bonds between adjacent thymine residues at carbon positions 5 and 6. If located within a gene, this will interfere with replication and transcription unless repaired.

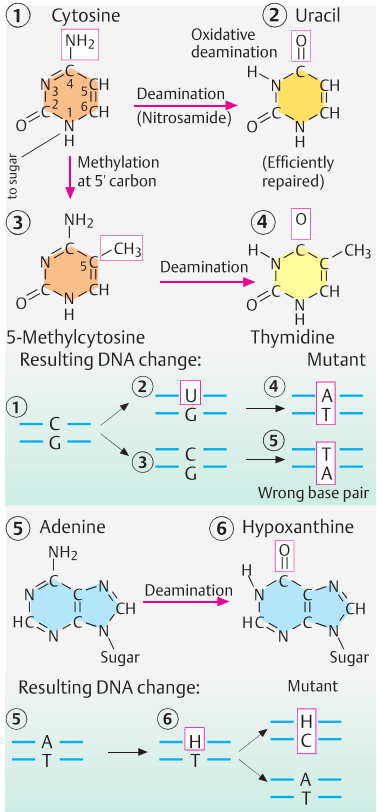
Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015: 268

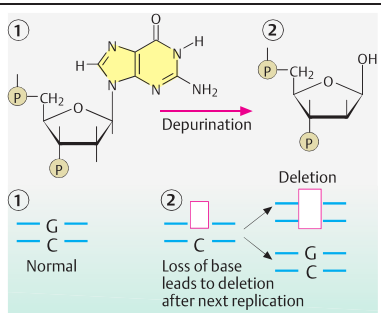
Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

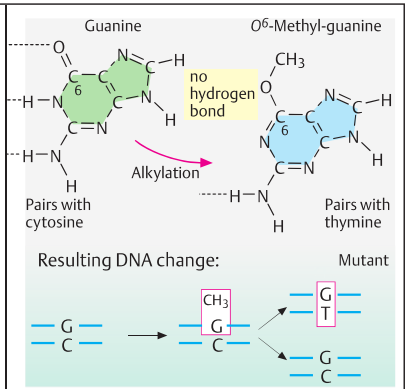
Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011



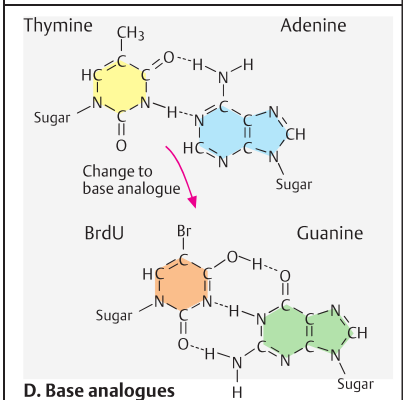
A. Deamination and methylation



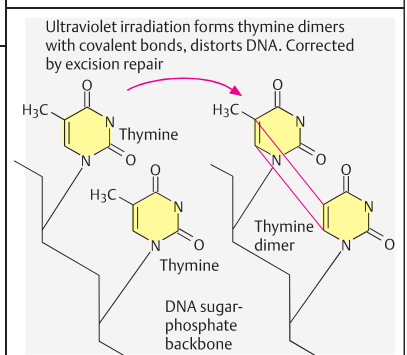
B. Depurination



C. Alkylation of guanine



D. Base analogues



E. UV-light-induced thymine dimers

Mutations Due to Errors in Replication

Errors occur frequently during replication. If an error goes undetected or is not repaired, it will result in different types of mutations. Errors in replication occur at a rate of approximately 1 in 10^5. Proofreading and repair system (see the next page) mechanisms reduce this rate to approximately 1 in 10^7 to 1 in 10^9. An important mechanism is replication slippage in regions with repeated nucleotide sequences.

A. Consequences of errors in replication

When an error in replication occurs before the next cell division, it might, for example, result in a cytosine (C) being incorporated instead of an adenine (A) at the fifth base pair, as shown here. If the error remains undetected, the next (second) division will result in a mutant molecule containing a CG instead of an AT pair at this position (blue field). This mutation will be perpetuated to all daughter cells.

B. Replication slippage

A different class of mutations does not involve an alteration of individual nucleotides, but results from incorrect alignment between allelic and nonallelic DNA sequences during replication. When the template strand contains short tandem repeats, for example, CA repeats as in microsatellites (see DNA polymorphisms), the newly replicated strand and the template strand may shift their positions relative to each other (microsatellite instability). With replication or polymerase slippage, leading to incorrect pairing of repeats, some repeats are copied twice and others not at all, depending on the direction of the shift. Thus, one can distinguish forward slippage and backward slippage in relation to the newly replicated strand. Backward slippage of the new strand results in the addition (insertion) of nucleotides to the new strand. Forward slippage of the new DNA strand results in the loss (deletion) of nucleotides from the new DNA.

C. Functional consequences of mutations

Aside from their molecular type, mutations can be classified according to their functional consequences (molecular pathology). A principal goal is to understand the relationship between the genotype and the phenotype. This is referred to as *genotype/phenotype relation* (often incorrectly named *correlation*).

The main class is loss of function of a normal allele. Here the gene product is reduced or has no function. When both alleles are necessary for normal function, but one is inactivated by a mutation, the result is called *haploinsufficiency*. The opposite is an undesirable functional effect of a new gene product resulting from a mutation; this is called a *dominant negative effect*. Overexpression of a normal gene product with undesirable effects is caused by a gain-of-function mutation. An epigenetic change is caused by an alteration other than that of the DNA sequence. Dynamic mutations result from the abnormal expansion of nucleotide repeats (see p. 84).

Medical relevance

Mutations cause more than 3,000 known individually defined diseases (see OMIM).

Microsatellite instability is a characteristic feature of hereditary nonpolyposis cancer of the colon (HNPCC). HNPCC is a genetically heterogeneous group of causes due to mutations in at least seven genes (OMIM 120435, 614331). Approximately 15% of all colorectal, gastric, and endometrial carcinomas show microsatellite instability. Replication slippage has to be distinguished from unequal crossing-over during meiosis. This is the result of recombination between adjacent sequences of homologous chromosomes.

Further Reading

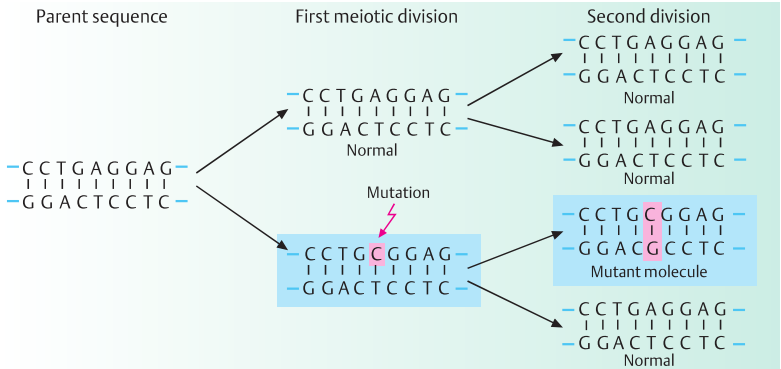
Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Krebs JE, et al. Lewin's. Genes XI. Sudbury: Bartlett & Jones, 2013

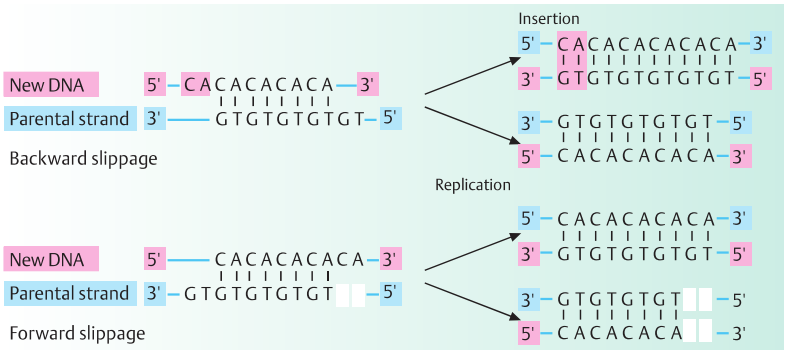
OMIM. Online Mendelian Inheritance of Man. Available at http://www.ncbi.nlm.nih.gov/omim

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015



A. Errors in replication



B. Replication slippage

	Two alleles	Genetic mechanisms involved
1. Normal		
2. Loss of function		Type of change causing loss of function:
Haploinsufficiency		• Frameshift resulting from deletion /insertion
Dominant negative effect		• Premature termination codon • Nonsense-mediated mRNA decay • Alteration of splice site • Interference with normal gene product
3. Gain of function		Undesirable overexpression of an allele
4. Epigenetic change		Change in DNA methylation pattern
5. Dynamic mutations		Expansion of nucleotide triplet repeats

C. Functional consequences of mutations

DNA Repair Systems

DNA damage can be recognized and repaired by multistep DNA repair pathways, each devoted to specific types of lesions and involving many types of repair proteins. Three major types of DNA repair are (i) base-excision repair, (ii) mismatch repair, and (iii) repair of double-strand breaks. In addition, single-strand break repair and nucleotide excision repair (for repair of bulky lesions distorting the DNA helix) exist.

A. Base-excision repair

This system removes 27 to 29 nucleotides of the damaged strand (~12 or 13 in prokaryotes). It recognizes a distorted damaged DNA strand, for example, by thymine–thymine dimers induced by ultraviolet light. Three proteins, XPA, XPB, and XPC in human cells (UvrA, UvrB, and UvrC in prokaryotes), detect the damaged site and form a repair protein complex. XP-C and XP-G endonucleases cleave the damaged strand at two sites. DNA repair synthesis by the poly(ADP-ribose) polymerase 1 restores the missing stretch, and a DNA ligase closes the gap.

B. Mismatch repair

Mismatch repair is an excision resynthesis system initiated at a DNA break 100 to 1,000 nucleotides away from the mismatch. It corrects errors of replication (see the previous page). The most important mismatch repair proteins in humans are MSH1, MSH2, MSH3, and MSH6; MLH1 (homologous to MutH, MutL, and MutS in bacteria); and PMS2 (postmeiotic segregation 2). MSH2 and MLH1 bind to mismatched base pairs, whereas others cleave DNA and remove the strand with erroneous bases. DNA polymerase III replaces the damaged strand.

C. Nonhomologous end joining

Double-strand repair joins the two ends of the DNA following a double-strand break. The ends bind to specific dimeric proteins of 80 and 90 kDa (Ku80 and Ku90). Additional repair proteins form a repair complex (Artemis consisting of a DNA double-strand repair/V(D)J recombination protein that is mutated in human severe combined immune deficiency, a DNA-dependent protein kinase [DNA-PK], and XRCC4 [X-ray repair complementing defective repair]) and a protein binding p53 (see p. 318). In this type of repair, bases may be lost or added.

D. Double-strand repair by homologous recombination

Homologous recombination repairs double-strand breaks using several repair proteins: ATM, BRCA1, BRCA2, RAD51, and others such as the Fanconi anemia proteins (see p. 324). ATM is activated in response to DNA damage (**1**). This activates BRCA1 by phosphorylation at specific sites (**2**). Phosphorylated BRCA1 induces homologous recombination in cooperation with BRCA2 and RAD51, the mammalian homolog of the *Escherichia coli* RecA repair protein (**3**). Phosphorylated BRCA1 is also involved in transcription and transcription-coupled DNA repair (**4**). The identical sister chromatid is used to align the broken ends and insert missing information.

Medical relevance

Three important heterogeneous groups of excision repair diseases with increased risk for malignant tumors (see p. 332) result from mutations in genes encoding repair proteins: several types of xeroderma pigmentosum (OMIM 278700–278780), Cockayne syndrome types A and B (OMIM 216400, 133540), and trichothiodystrophy (OMIM 601675).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Cleaver JE, et al. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. Nat Rev Genet 2009; 10(11): 756–768

Fishel R. Mismatch repair. J Biol Chem 2015; 290(44): 26395–26403

Hoeijmakers JHJ. DNA damage, aging, and cancer. N Engl J Med 2009; 361(15): 1475–1485

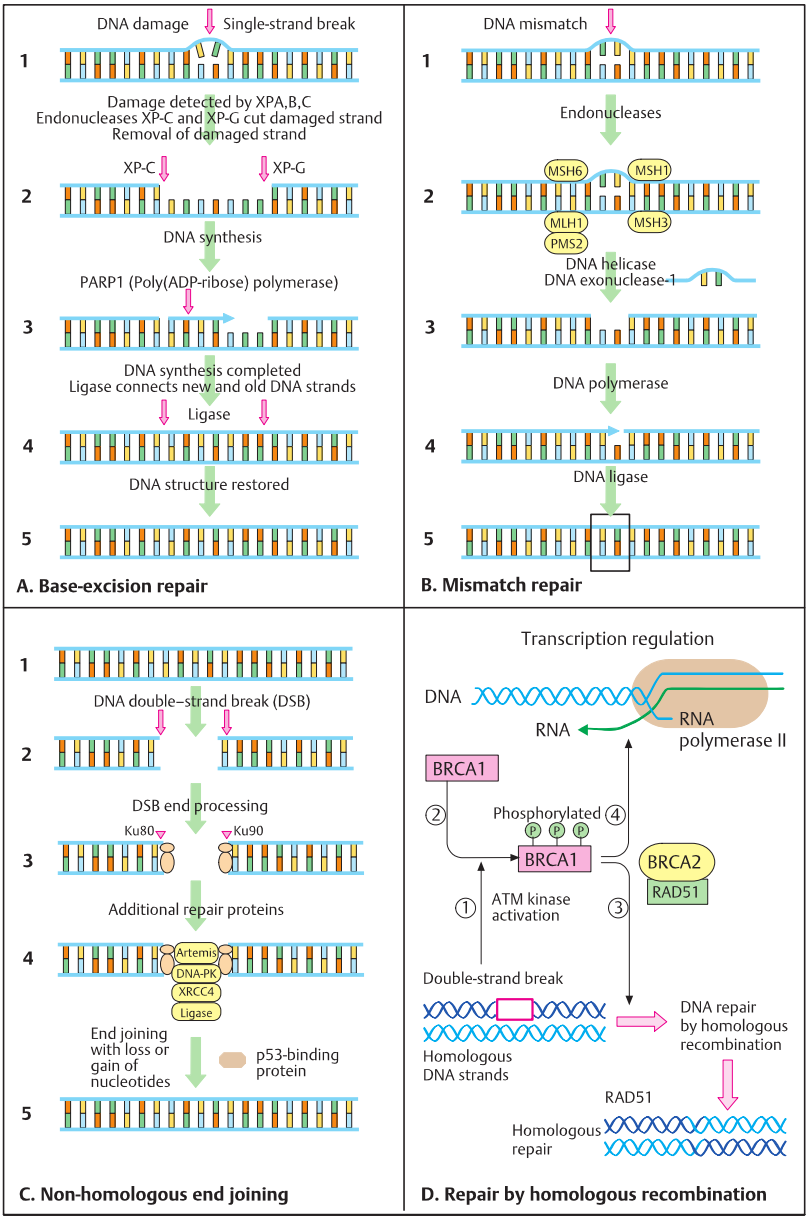
Liu J, et al. Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. Nature 2016; 539(7630): 583–587

Moscariello M, et al. Role for Artemis nuclease in the repair of radiation-induced DNA double strand breaks by alternative end joining. DNA Repair (Amst) 2015; 31: 29–40

O'Driscoll M, et al. The role of double-strand break repair - insights from human genetics. Nat Rev Genet 2006; 7(1): 45–54

OMIM. Online Mendelian Inheritance of Man. Available at http://www.ncbi.nlm.nih.gov/omim

Venkitaraman AR. Breast cancer genes and DNA repair. Science 1999; 286(5442): 1100–1102


DNA damage
Single-strand break
1
Damage detected by XPA,B,C
Endonucleases XP-C and XP-G cut damaged strand
Removal of damaged strand
XP-C
XP-G
2
DNA synthesis
PARP1 (Poly(ADP-ribose) polymerase)
3
DNA synthesis completed
Ligase connects new and old DNA strands
Ligase
4
DNA structure restored
5
A. Base-excision repair
DNA mismatch
1
Endonucleases
MSH6
MSH1
2
MLH1
MSH3
PMS2
DNA helicase
DNA exonuclease-1
3
DNA polymerase
4
DNA ligase
5
B. Mismatch repair
1
DNA double–strand break (DSB)
2
DSB end processing
Ku80
Ku90
3
Additional repair proteins
4
Artemis
DNA-PK
XRCC4
Ligase
End joining with loss or gain of nucleotides
p53-binding protein
5
C. Non-homologous end joining
Transcription regulation
DNA
RNA
RNA polymerase II
BRCA1
②
Phosphorylated ④
P
P
P
BRCA1
BRCA2
RAD51
①
ATM kinase activation
③
Double-strand break
DNA repair by homologous recombination
Homologous DNA strands
RAD51
Homologous repair
D. Repair by homologous recombination

Transposition

Transposition refers to parts of DNA that can change from one position to another, collectively called *transposons* (Tn) or *transposable elements*, first described by Barbara McClintock in 1951. A specific enzyme encoded by the transposon, transposase, mediates the insertion at a new target DNA site. Three examples of the different classes of transposons are presented. Transposons are a major source of genetic variation and play an important role in the evolution of genomes.

A. Insertion sequences and transposons (Tn)

The host DNA contains a target site of approximately 4 to 10 base pairs (bp) (**1**). The selection of the target site of the host DNA is either random or selective for particular sites. The insertion sequence (IS) consists of approximately 700 to 1,500 bp, depending on the particular class. Its transposase gene encodes the enzyme responsible for transposition. It is flanked by inverted repeats of approximately 9 bp at both ends. This is a characteristic feature of IS transposition. The IS inserts itself at the target site by means of the transposase activity (**2**). Transposons (Tn) may contain other genes, such as those for antibiotic resistance, and have direct (**3**) or inverted (**4**) repeats at either end. Direct repeats are identical or closely related sequences oriented in the same direction. Inverted repeats are oriented in opposite directions.

B. Replicative and nonreplicative transposition

In replicative transposition (**1**), the donor transposon remains in place and creates a new copy of itself, and this inserts into a recipient site elsewhere. This mechanism leads to an increase in the number of copies of the transposon in the genome. It involves two enzymatic activities: a transposase, acting on the ends of the original transposon, and resolvase, acting on the duplicated copies. In nonreplicative transposition (**2**), the transposing element itself moves as a physical entity directly to another site.

C. Transposition of retroelements

Transposition is used by viruses with an RNA genome (RNA retroviruses) to insert themselves into the host DNA genome (retrotransposition) by reverse transcription from RNA to DNA (reverse transcriptase). This requires synthesis of an RNA copy of the inserted retroelement. Retroviruses, including the human immunodeficiency virus (HIV) and RNA tumor viruses, are important retroelements. The first step in retrotransposition is the synthesis of an RNA copy of the inserted retroelement, followed by reverse transcription up to the polyadenylation sequence in the 3′ long terminal repeat (LTR). Three important classes of mammalian transposons that undergo or have undergone retrotransposition through an RNA intermediary are shown. Endogenous retroviruses (**1**) are sequences that resemble retroviruses but cannot infect new cells and are restricted to one genome. Nonviral retrotransposons (**2**) lack LTRs and usually other parts of retroviruses. Both types contain reverse transcriptase and are therefore capable of independent transposition. Processed pseudogenes (**3**) or retropseudogenes lack reverse transcriptase and cannot transpose independently. They contain two groups: (i) low copy number of processed pseudogenes transcribed by RNA polymerase II and (ii) high copy number of mammalian SINE sequences, such as human Alu and the mouse B1 repeat families. One in 600 mutations is estimated to arise from retrotransposon-mediated insertion.

Medical relevance

Transposition may be the cause of structural disease-causing changes in a gene.

Further Reading

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

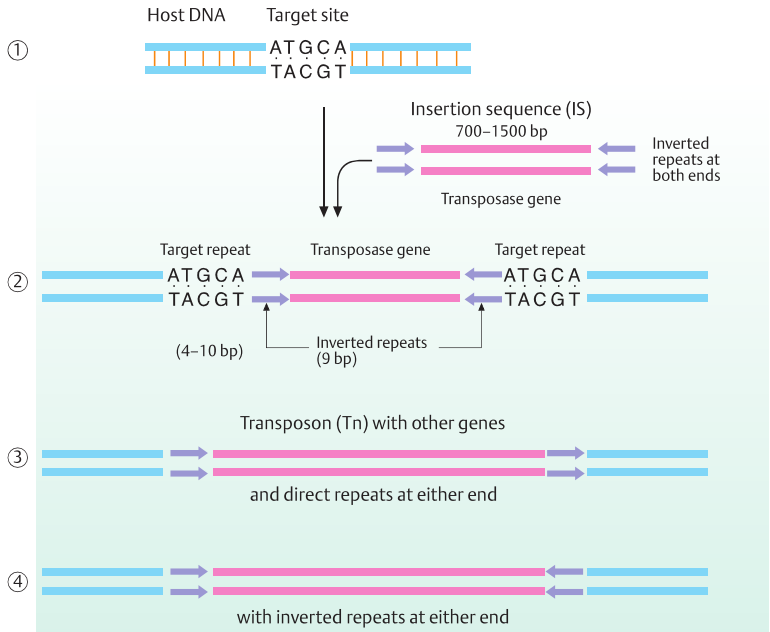
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Mills RE, et al. Which transposable elements are active in the human genome? Trends Genet 2007; 23(4): 183–191

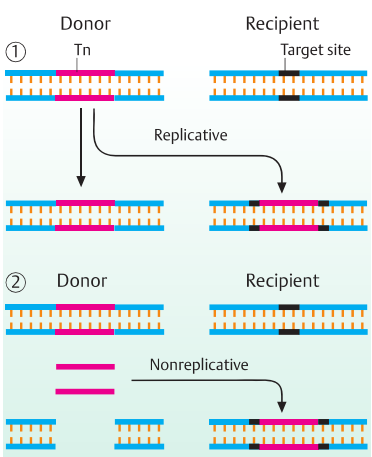
Prak ET, et al. Mobile elements and the human genome. Nat Rev Genet 2000; 1(2): 134–144

Reilly MT, et al. The role of transposable elements in health and diseases of the central nervous system. J Neurosci 2013; 33(45): 17577–17586

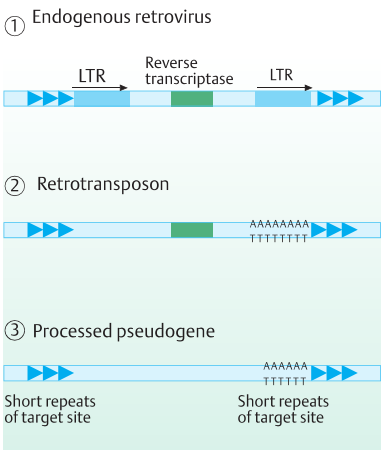
Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011



A. Insertion sequences (IS) and transposons (Tn)



B. Replicative and nonreplicative transposition



C. Transposition of retroelements

Trinucleotide Repeat Expansion

Short nucleotide repeats may expand abnormally and cause a variety of different diseases, mainly of the nervous system. They constitute a new class of mutations discovered in 1991. These are pathogenic microsatellite variants that may become unstable (*unstable mutations*). The human genome contains numerous short tandem repeats of three nucleotides (trinucleotides or triplets) or more. If expanded abnormally within or near certain genes, they interfere with gene expression (*trinucleotide expansion disorders*). Normally, triplets usually occur in groups of 5 to 35 repeats, but this varies depending on the site. Although usually transmitted stably, they can become unstable and expand to pathological lengths. Once the normal length has expanded, the number of repeats tends to increase when passed through the germline. This causes an earlier onset of the disease than in the preceding generations, an observation called *anticipation*.

A. Different types of trinucleotide repeats and their expansions

Trinucleotide repeats can be distinguished according to their location with respect to a gene. Very long expansions occur in introns (**1**). The increase in the number of these repeats can be drastic, up to 1,000 or more repeats. The first stages of expansion do not usually lead to clinical signs, but predispose to increasing expansion of the repeat in the offspring of a carrier (premutation). Within the coding regions, in exons, expansions are more moderate (**2**). However, their effect is dramatic, as seen in several severe neurologic diseases, because they result in expanded glutamine (CAG) tracts.

B. Unstable trinucleotide repeats in different diseases

Disorders caused by pathological expansion of trinucleotide repeats are classified according to the type of trinucleotide repeat, that is, the sequence of the three nucleotides, their location with respect to the gene involved, and their clinical features. All involve the central or the peripheral nervous system. Type I trinucleotide diseases are characterized by CAG trinucleotide expansions within the coding regions of different genes. The triplet CAG encoding glutamine usually has approximately 20 CAG repeats. Thus, approximately 20 glutamine residues occur in the gene product. In the disease state, the number of glutamine residues is greatly increased in the protein. Hence, the diseases are collectively referred to as polyglutamine disorders.

Type II trinucleotide diseases are characterized by expansions of CTG, GAA, GCC, or CGG trinucleotides within a noncoding region of the gene involved at the 5′ UTR (untranslated region of exon 1 [CGG in fragile X syndrome type A, FRAXA]), at the 3′ end (CGG in FRAXE; CTG in myotonic dystrophy), or in an intron (GAA in Friedreich ataxia). A brief review of these disorders is given on page 364.

C. Principle of laboratory diagnosis

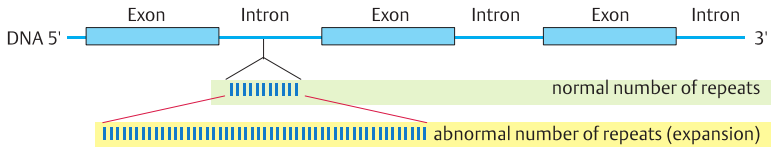
The laboratory diagnosis compares the sizes of the trinucleotide repeats in the two alleles of the gene by Southern blot hybridization (see p. 70). The schematic figure shows 11 lanes, each representing one individual: normal controls (lanes 1–3) and patients with Huntington chorea (lanes 4–7 and 10). A family is represented in lanes 7–11: an affected father (lane 7), an affected son (lane 10), an unaffected mother (lane 11), and two unaffected children (a son, lane 8; and a daughter, lane 9). Size markers are shown at the left. Each lane represents a polyacrylamide gel, and the (CAG) repeat of the Huntington locus amplified by polymerase chain reaction is shown as a band of defined size. The two alleles are shown for each individual. In the affected individuals, the band representing the abnormal allele is located above the threshold in the expanded region (in practice, the bands may be blurred because the exact repeat size varies in DNA from different cells).

Further Reading

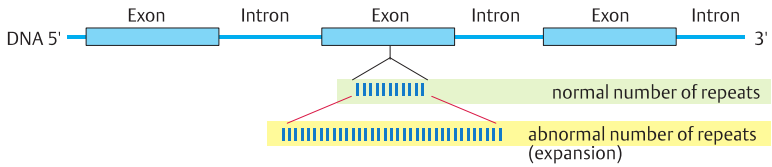
Antonarakis SE, et al. Expansion/copy number variation of trinucleotide (and other) repeat sequences. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013: 8–9 (e-book)

Orr HT, et al. Trinucleotide repeat disorders. Annu Rev Neurosci 2007; 30: 575–621

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015

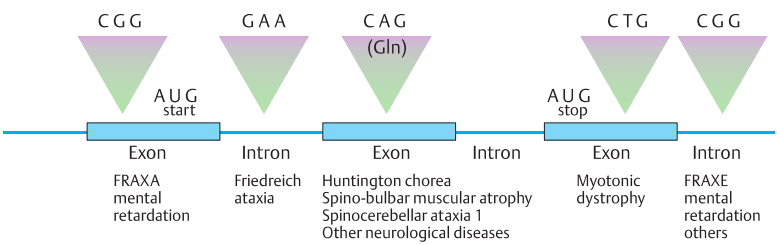


1. Very large expansions of repeats outside coding sequences

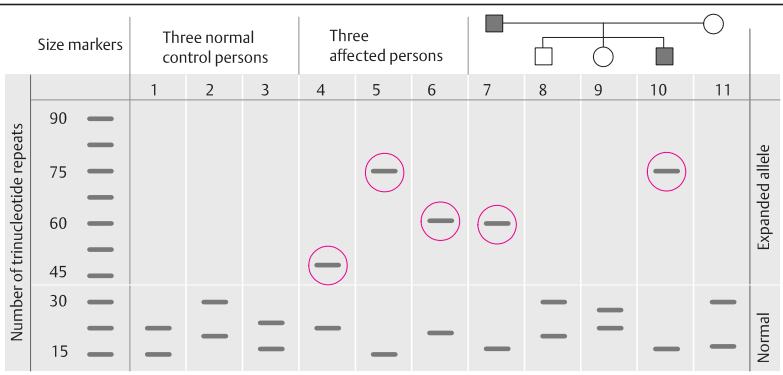


2. Modest expansion of CAG repeats within coding sequences

A. Different types of trinucleotide repeat expansion



B. Unstable trinucleotide repeats in different diseases



C. Principle of laboratory diagnosis of unstable trinucleotide repeats leading to expansion

Cell Communication

Multicellular organisms use extensive systems by which cells communicate through a vast variety of extracellular signal proteins that mediate specific intra-cell responses, some over a long distance. These proteins include extracellular signal molecules, cell surface receptors, intracellular receptors, and intracellular signal molecules that transmit signals. These regulate growth (e.g., embryonic development), cell differentiation into one of the approximately 200 different cell types, the cell cycle, and other cellular functions.

A. Principle of signal transduction

The transduction of a signal elicits a cell-specific effect. This involves several types of interacting proteins. A cell membrane-bound receptor, consisting of an extracellular and an intracellular portion (called *domains*), responds to a signal molecule. The specificity of the response is achieved by the binding of specific signal molecule (called a *ligand*) to the extracellular domain of the receptor. This, in turn, activates a series of further signaling proteins downstream. Generally, one activated protein activates the next by a specific biochemical reaction, called a *signaling cascade*. Only two are shown here (designated signaling proteins 1 and 2), but quite often many more are involved. The final steps of a signaling cascade reach the target protein and elicit the desired cellular response. Extracellular signal molecules typically act at very low concentrations, at approximately 10^{-8} molar concentration.

B. Signaling between cells

Different types of cell signals exist. Often they form a signal pathway that transmits the signal through several relays to the target. A signal molecule may remain attached to the surface of the signaling cell when binding to the target cell (**1**, *contact-dependent signaling*). This type of signaling is common in embryonic development and in the immune system. In (**2**), *paracrine signaling*, the signal molecule is secreted and released into the extracellular space. Here it acts over a short distance on cells in the neighborhood. Many growth and differentiation factors act in this manner. Some important long-distance signaling is mediated through hormones in the blood circulation (**3**, *endocrine signaling*). In this type of signaling, specialized cells, called *endocrine cells*, secrete a substance, called a *hormone*, into the bloodstream. From there, it can reach the target cells at a distance in another part of the body. *Synaptic signaling* refers to nerve cells or the junction of nerve and muscle cells (**4**). In this case, a specialized cell, a nerve cell (neuron), sends electrical impulses along a cellular extension, the *axon*, which can be quite long. At the end of the axon, a chemical signal, called a *neurotransmitter*, is secreted at the junction (the synapse) between the signaling cell (the neuron) and the postsynaptic target cell. In some cases, the same types of signaling molecules are used in paracrine, endocrine, and synaptic signaling, but in different contexts of selectivity. (Figure adapted from Lodish et al., 2016.)

The term *hormone* is derived from the Greek word *hormon*, meaning to 'spur on'. It was first used in 1904 by William Bayliss and Ernest Starling to describe the action of a secreted molecule. Five major classes of hormones can be defined: (i) amino acid derivatives (e.g., catecholamine, dopamine, thyroxine), (ii) small neuropeptides (e.g., thyrotropin-releasing hormone, somatostatin, vasopressin), (iii) proteins (e.g., insulin, luteinizing hormone), (iv) steroid hormones derived from cholesterol (e.g., cortisol, sex hormones), and (v) vitamin derivatives (e.g., retinoids [vitamin A], peptide growth hormones).

Medical relevance

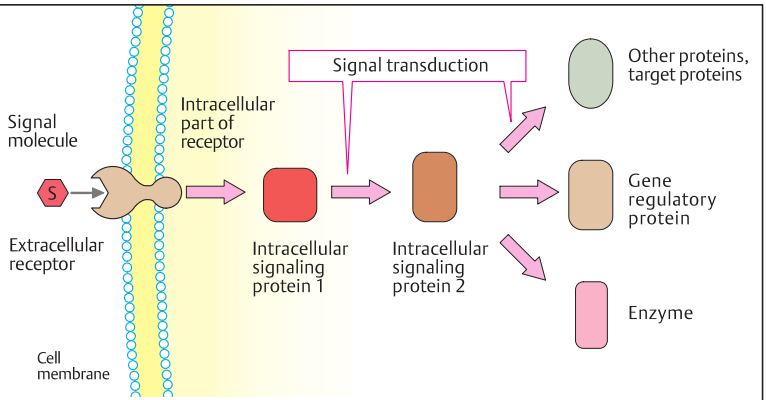
Mutations in the genes encoding proteins involved in signal transduction cause a vast array of different human genetic disorders.

Further Reading

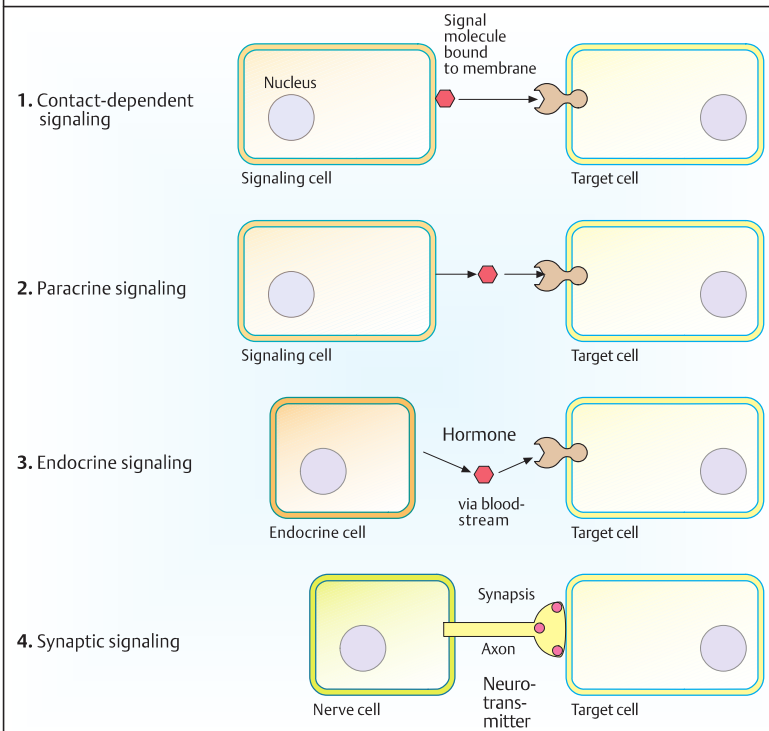
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Jameson JL. Principles of endocrinology. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 2866–2875

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016



A. Principle of signal transduction



B. Signaling between cells

Haploid and Diploid Yeast Cells

Budding yeast, the common baker's yeast (*Saccharomyces cerevisiae*), is an important model organism that is used widely in genetics. It is easy to grow and manipulate. It can differentiate into three distinct cell types: two haploid and one diploid. It is a single-celled eukaryotic fungus with a genome of individual linear chromosomes enclosed in a nucleus. It contains cytoplasmic organelles such as endoplasmic reticulum, Golgi apparatus, mitochondria, peroxisomes, and a vacuole analogous to a lysosome. Approximately 40 different types of yeast are known. Baker's yeast consists of oval cells approximately 5 µm in diameter. Under good nutritional conditions, a cell can divide by budding every 90 minutes. Fission yeast, *Schizosaccharomyces pombe*, has rod-shaped cells that divide by elongation at the ends.

The haploid genome of *S. cerevisiae* contains approximately 6,200 genes in 1.4×10^7 DNA base pairs distributed in 16 chromosomes (Goffeau et al., 1996). The genes are involved in the following functions: cell structure, 250 genes (4%); DNA metabolism, 175 genes (3%); transcription and translation, 750 genes (13%); energy production and storage, 175 genes (3%); biochemical metabolism, 650 genes (11%); and transport, 250 genes (4%). The *S. cerevisiae* genome is very compact compared with other eukaryotic genomes, with approximately one gene every 2 kilobases. Nearly half of the human proteins known to be defective in hereditary disease have amino acid similarities to a yeast protein.

A. Yeast life cycle

The life cycle of yeast passes through a haploid or a diploid phase. Haploid cells of opposite types can fuse (mate) to form a diploid cell. Haploid cells are of one of two possible mating types, called *a* and α. The mating is mediated by a small secreted polypeptide called a *pheromone* or *mating factor*. A cell-surface receptor recognizes the pheromone secreted by cells of the opposite type, that is, **a** cell receptors bind only **α** factor, and **α** cell receptors bind only **a** factor. Mating and subsequent mitotic divisions occur in favorable conditions for growth. In starvation conditions, a diploid yeast cell undergoes meiosis and forms four haploid spores (*sporulation*): two of type **a** and two of type **α**.

B. Switch of mating type

The switch of mating type (mating-type conversion) is initiated by a double-strand break in the DNA at the *MAT* locus (recipient) and may involve the boundary to either of the flanking donor loci (*HMR* or *HML*). This is mediated by a HO endonuclease through site-specific DNA cleavage.

C. Cassette model for mating-type switch

Mating-type switch is regulated at three gene loci near the centromere (cen) of chromosome III of *S. cerevisiae*. The central locus is *MAT* (mating-type locus), which is flanked by loci *HML*α (left) and *HMR*a (right). Only the *MAT* locus is active and transcribed into mRNA. Transcription factors regulate other genes responsible for the **a** or α phenotype. The *HML*α and *HMR*a loci are repressed (silenced). DNA sequences from either the *HML*α or the *HML*a locus are transferred into the *MAT* locus once during each cell generation by a specific recombination event called *gene conversion*. The presence of *HMR*α sequences at the MAT locus determines the phenotype of the **a** cell. When *HML*α sequences are transferred (switched to an α cassette), the phenotype is switched to **α**. A gene is repressed when it is placed near the yeast mating-type silencer.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Botstein D, et al. Yeast as a model organism. Science 1997; 277(5330): 1259–1260

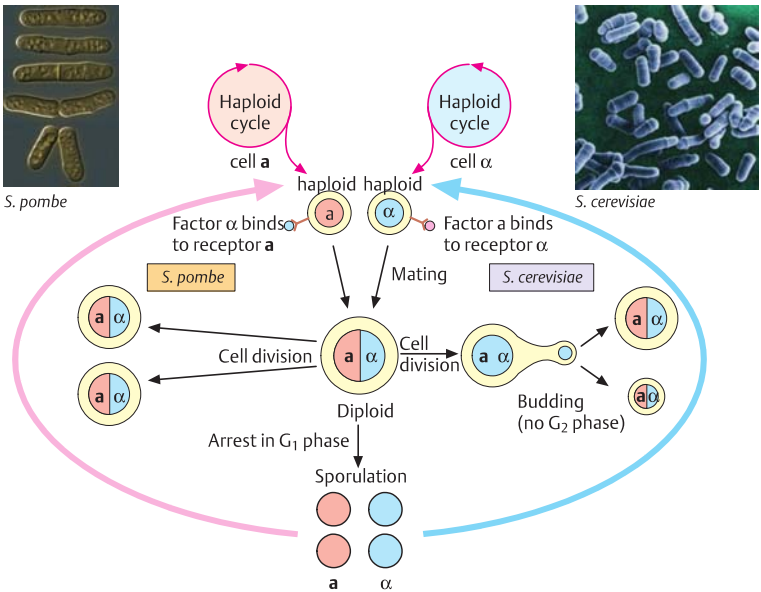
Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Goffeau A, et al. Life with 6000 genes. Science 1996; 274(5287): 546, 563–567

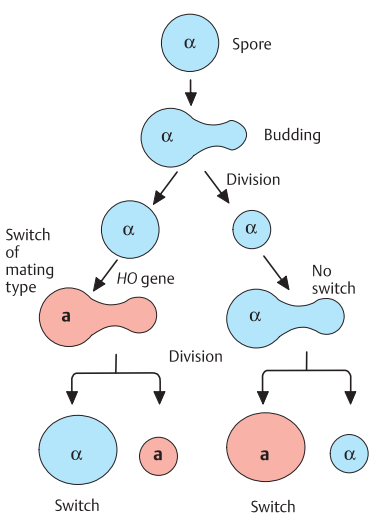
Haber JE. A locus control region regulates yeast recombination. Trends Genet 1998; 14(8): 317–321

Lewin B. Chromosomes IX. Sudbury: Bartlett & Jones, 2008

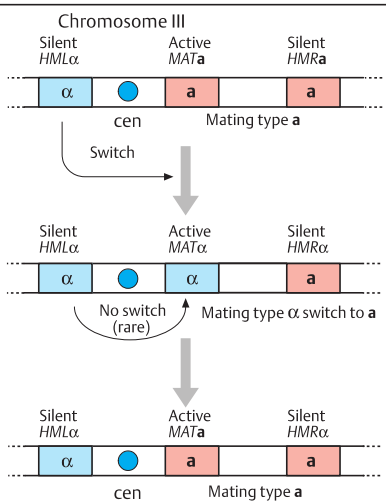
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016



A. Yeast life cycle through a haploid and a diploid phase



B. Switch of mating type



C. Cassette model for mating–type switch

Cell Cycle Control

Cell division results in two identical daughter cells with an identical set of chromosomes. The cell cycle, as first defined by Howard and Pelc in 1953, has two main phases: interphase and mitosis. An elaborate set of controlling macromolecules guides the cell through its cycle. Here, control systems can detect and eliminate errors, and abandon defective cells.

A. Yeast cells

Eukaryote cells have a similar cell cycle and its control. Genetic studies in yeast cells have provided important insights into cell cycle control. Baker's yeast *(Saccharomyces cerevisiae)* and fission yeast (*Schizosaccharomyces pombe*) have cell cycle control mechanisms that are similar to those of higher eukaryotes.

B. Cell division cycle models in yeast

Mitotic division in budding yeast (baker's yeast) results in one large and one small daughter cell. Since a microtubule mitotic spindle forms very early during the S phase, there is practically no G_2 phase (**1**). In contrast, fission yeast (*S. pombe*) forms a mitotic spindle at the end of the G_2 phase and then proceeds to mitosis to form two daughter cells of equal size (**2**). Unlike in vertebrate cells, the nuclear envelope remains intact during mitosis. An important regulator of yeast cell division is cdc2 (cell division cycle 2) protein. The absence of cdc2 activity (cdc2 mutants) in *S. pombe* results in cycle delay and prevents entry into mitosis (**3**). Thus, the result is a cell that is too large and only has one nucleus. Increased activity of cdc2 (dominant mutant cdc^D) results in premature mitosis and cells that are too small (wee phenotype, from the Scottish word for small).

C. Cell cycle control systems

The cell cycle is marked by four phases, G_1, M, S, and G_2. G_1 (gap1) is followed by M (mitosis) wherein S (synthesis of DNA occurs). When mitosis is completed the cell reaches G_2. G_1, S, and G_2 are the interphase. Several different proteins and growth factors initiate cell division. Other systems detect DNA or cell damage and induce various repair systems (see p. 80). The eukaryotic cell cycle is driven by cell cycle "engines," which are a set of interacting proteins known as the cyclin-dependent kinases (Cdks). An important member of this family of proteins is cdc2 (also called *Cdk1*). Cell cycle arrest is initiated by a variety of genes after DNA damage has been detected. The p53 protein (see p. 318) plays a leading role in detecting DNA damage and subsequent cell cycle arrest.

In early G_1 phase, cdc2 is inactive. It is activated in late G_1 by associating with G_1 cyclins, such as cyclin E. Once the cell has passed the G_1 checkpoint, cyclin E is degraded and the cell enters the S phase. This is initiated, among many other activities, by cyclin A binding to Cdk2 and phosphorylation of the retinoblastoma (RB) protein (see p. 326). The cell can pass through the mitosis checkpoint provided no damage is present. Association of Cdc2 (Cdk1) with mitotic cyclins A and B activates and forms the mitosis-promoting factor. During mitosis, cyclins A and B are degraded, and an anaphase-promoting complex is formed (details not shown). When mitosis is completed, cdc2 is inactivated by the S phase inhibitor Sic1 in yeast. At the same time, the RB protein is dephosphorylated. Cells can progress to the next cell cycle stage only when feedback controls have ensured the integrity of the genome.

Medical relevance

Mutations in one of the many cell cycle controlling genes may result in different types of cancer.

Further Reading

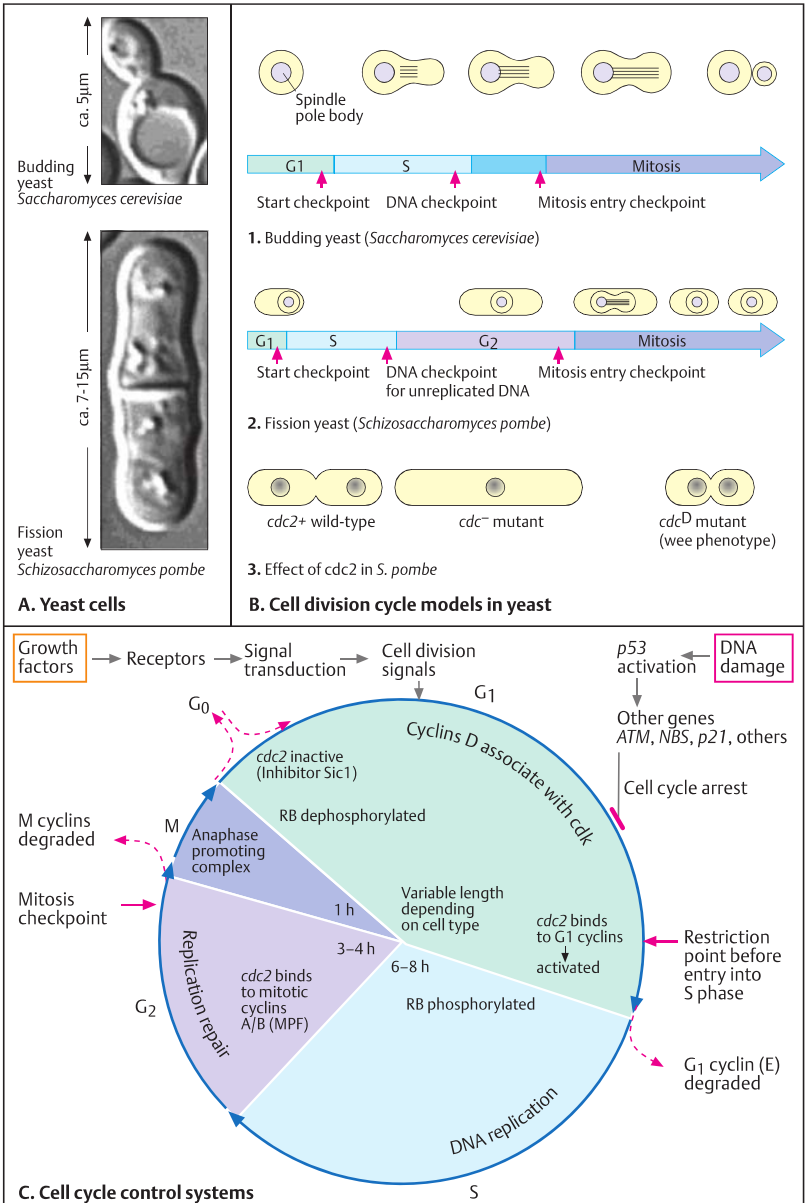
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Hartwell LH, Weinert TA. Checkpoints: controls that ensure the order of cell cycle events. Science 1989; 246(4930): 629–634

Howard A, Pelc S. Synthesis of deoxyribonucleic acid in normal and irradiated cells and its relation to chromosome breakage. Heredity 1953; 6: 261–273

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Nurse P. A long twentieth century of the cell cycle and beyond. Cell 2000; 100(1): 71–78


ca. 5µm
Budding yeast
Saccharomyces cerevisiae
ca. 7-15µm
Fission yeast
Schizosaccharomyces pombe
A. Yeast cells
Spindle pole body
G1
S
Mitosis
Start checkpoint
DNA checkpoint
Mitosis entry checkpoint
1. Budding yeast (Saccharomyces cerevisiae)
G1
S
G2
Mitosis
Start checkpoint
DNA checkpoint for unreplicated DNA
Mitosis entry checkpoint
2. Fission yeast (Schizosaccharomyces pombe)
cdc2+ wild-type
cdc⁻ mutant
cdcD mutant (wee phenotype)
3. Effect of cdc2 in S. pombe
B. Cell division cycle models in yeast
Growth factors
Receptors
Signal transduction
Cell division signals
p53 activation
DNA damage
G0
G1
Cyclins D associate with cdk
cdc2 inactive (Inhibitor Sic1)
Other genes ATM, NBS, p21, others
Cell cycle arrest
M cyclins degraded
M
Anaphase promoting complex
RB dephosphorylated
Mitosis checkpoint
1 h
Variable length depending on cell type
cdc2 binds to G1 cyclins
activated
Restriction point before entry into S phase
3–4 h
6–8 h
Replication repair
cdc2 binds to mitotic cyclins A/B (MPF)
G2
RB phosphorylated
G1 cyclin (E) degraded
DNA replication
S
C. Cell cycle control systems

Cell Division: Mitosis

Mitosis is the process of cell division. This term, introduced by W. Flemming in 1882, is derived from the Greek word μίτος (mitos), meaning a thread. Threadlike structures in dividing cells were first observed by Flemming in 1879. In 1884, E. Strasburger coined the terms *prophase, metaphase*, and *anaphase* for the different stages of cell division. During mitosis, the newly replicated chromosomes are distributed equally into different parts of a cell and then separated into their own cell. This results in two genetically identical daughter cells.

A. Mitosis

A cell divides in four consecutive phases. In eukaryotic cells, each cell division begins with a phase of DNA synthesis, which lasts approximately 8 hours (*S phase*). This is followed by the *G_2 phase* of approximately 4 hours (gap 2) until the onset of mitosis (M). Mitosis in eukaryotic cells lasts approximately 1 hour. It follows the interphase (G_1) of extremely variable duration. Cells that no longer divide are in the G_0 phase.

During the transition from interphase to mitosis, the chromosomes become visible as elongated threads; this first phase of mitosis is called *prophase*. The chromosomes contract during late *prophase* to become thicker and shorter (chromosomal condensation). In late prophase, the nuclear membrane disappears and *metaphase* begins. The chromosomes become arranged on the equatorial plate, but homologous chromosomes do not pair. In late metaphase, during the transition into *anaphase*, the chromosomes divide at the centromere region. The two chromatids of each chromosome migrate to opposite poles, and *telophase* begins with the formation of the nuclear membranes. Finally, the cytoplasm also divides (cytokinesis). In early interphase, the individual chromosomal structures become invisible in the cell nucleus.

B. Metaphase chromosome

In 1888, Waldeyer coined the term *chromosome* for the stainable threadlike structures visible during mitosis. A metaphase chromosome consists of two chromatids (*sister chromatids*) and the centromere, which holds them together. The regions at both ends of the chromosome are the telomeres. The point of attachment to the mitotic spindle fibers is the kinetochore.

C. Role of condensins

The progressive compaction of chromosomes entering mitosis is called *chromosome condensation*. A mitotic chromosome is approximately 50 times shorter during interphase. Therefore, it becomes microscopically visible. Chromosome condensation is mediated by proteins called *condensins*. These consist of five subunits (not shown). Condensins can be visualized along a mitotic chromosome.

When chromosomes are duplicated in S phase, the two copies of each chromosome remain tightly bound together as sister chromatids. They are held together by multiunit proteins called *cohesins*. Cohesins consist of four subunits and are structurally related to condensins. They regulate the separation of sister chromatids. Mutations in fission yeast cohesins interfere with mitosis.)

Medical relevance

A mutation in any of the five genes encoding condensin subunits results in a severe growth and malformation syndrome: Roberts syndrome (OMIM 268300; Vega et al., 2005). For disorders resulting from faulty cohesins, see p. 246.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Karsenti E, et al. The mitotic spindle: a self-made machine. Science 2001; 294(5542): 543–547

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

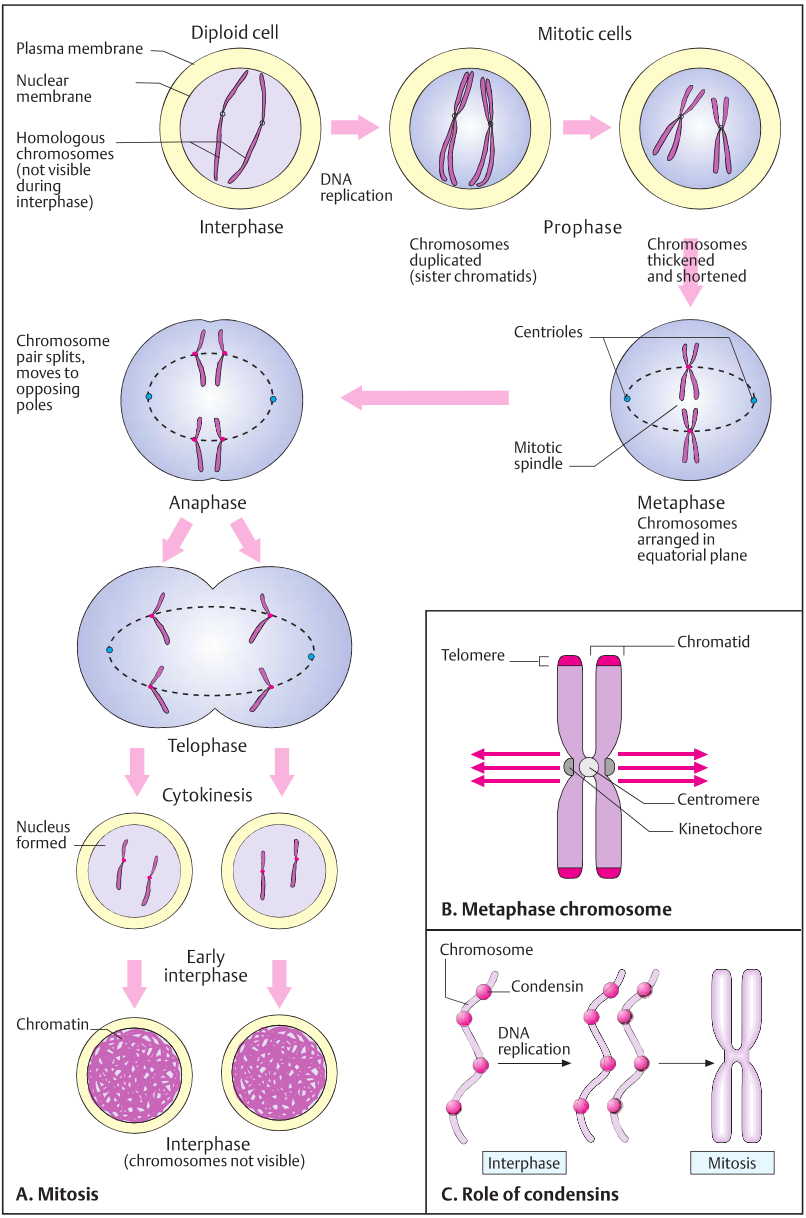
Nurse P. The incredible life and times of biological cells. Science 2000; 289(5485): 1711–1716

Rieder CL, Khodjakov A. Mitosis through the microscope: advances in seeing inside live dividing cells. Science 2003; 300(5616): 91–96

Tsukahara T, et al. Phosphorylation of the CPC by Cdk1 promotes chromosome bi-orientation. Nature 2010; 467(7316): 719–723

Uhlmann F, et al. Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. Nature 1999; 400(6739): 37–42

Vega H, et al. Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet 2005; 37(5): 468–470


Diploid cell
Plasma membrane
Nuclear membrane
Homologous chromosomes (not visible during interphase)
Interphase
DNA replication
Mitotic cells
Chromosomes duplicated (sister chromatids)
Prophase
Chromosomes thickened and shortened
Centrioles
Mitotic spindle
Metaphase
Chromosomes arranged in equatorial plane
Chromosome pair splits, moves to opposing poles
Anaphase
Telophase
Cytokinesis
Nucleus formed
Early interphase
Chromatin
Interphase
(chromosomes not visible)
A. Mitosis
Telomere
Chromatid
Centromere
Kinetochore
B. Metaphase chromosome
Chromosome
Condensin
DNA replication
Interphase
Mitosis
C. Role of condensins

Meiosis in Germ Cells

Meiosis (derived from μειόω: to reduce) is a special type of cell division that in most eukaryotes produces haploid germ cells (eggs and sperm cells). This term was introduced by Strasburger in 1884 and is derived from a Greek word meaning diminution (or maturation division). Meiosis consists of two nuclear divisions but only one round of DNA replication. As a result, the four daughter cells are haploid, that is, they contain only one chromosome of each pair.

Meiosis differs fundamentally from mitosis in genetic and cytological respects. First, homologous chromosomes pair at prophase of the first division. Second, exchanges between homologous chromosomes (*crossing-over*) occur regularly. As a result, all new chromosomes consist of segments of both maternal and paternal origin. The process of creating new combinations of genetic information is called *genetic recombination*. Third, the chromosome complement is reduced to half during the first cell division, meiosis I.

Meiosis is a complex cellular and biochemical process. The cytologically observable course of events and the genetic consequences do not correspond exactly in time. A genetic process occurring in one phase usually becomes visible cytologically at a later phase.

A. Meiosis I

A gamete-producing cell goes through two cell divisions at meiosis: meiosis I and meiosis II. The relevant genetic events—genetic recombination by means of crossing-over and reduction to the haploid chromosome complement—occur in meiosis I. Meiosis begins with DNA replication. At the beginning of prophase I, the chromosomes are duplicated. The pairing allows an exchange between homologous chromosomes (crossing-over), made possible by juxtapositioning homologous chromatids. At certain sites, a chiasma forms. As a result of crossing-over, chromosome material of maternal and paternal origin is exchanged between two chromatids of homologous chromosomes. After the homologous chromosomes migrate to opposite poles, the cell enters anaphase I.

B. Meiosis II

Meiosis II consists of longitudinal division of the duplicated chromosomes (chromatids) and a further cell division. Each daughter cell is haploid, as it contains one chromosome of a pair only. On each chromosome, recombinant and nonrecombinant sections can be identified. The genetic events relevant to these changes have occurred in the prophase of meiosis I (see next page).

During meiosis, several meiosis-specific proteins in the cohesion complex ensure correct homologous pairing and subsequent separation.

Medical relevance

The independent distribution of chromosomes (independent assortment) during meiosis explains the segregation (separation or splitting) of observable traits according to the rules of Mendelian inheritance (see p. 110).

Errors in the correct distribution of the chromosomes, called *nondisjunction*, result in gametes with an extra chromosome or a chromosome missing, and after fertilization, the zygote will have either three homologous chromosomes (*trisomy*) or only one chromosome (*monosomy*) (see chromsomal disorders, p. 156, 382).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Carpenter ATC. Chiasma function. Cell 1994; 77(7): 957–962

Kitajima TS, et al. Distinct cohesin complexes organize meiotic chromosome domains. Science 2003; 300(5622): 1152–1155

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

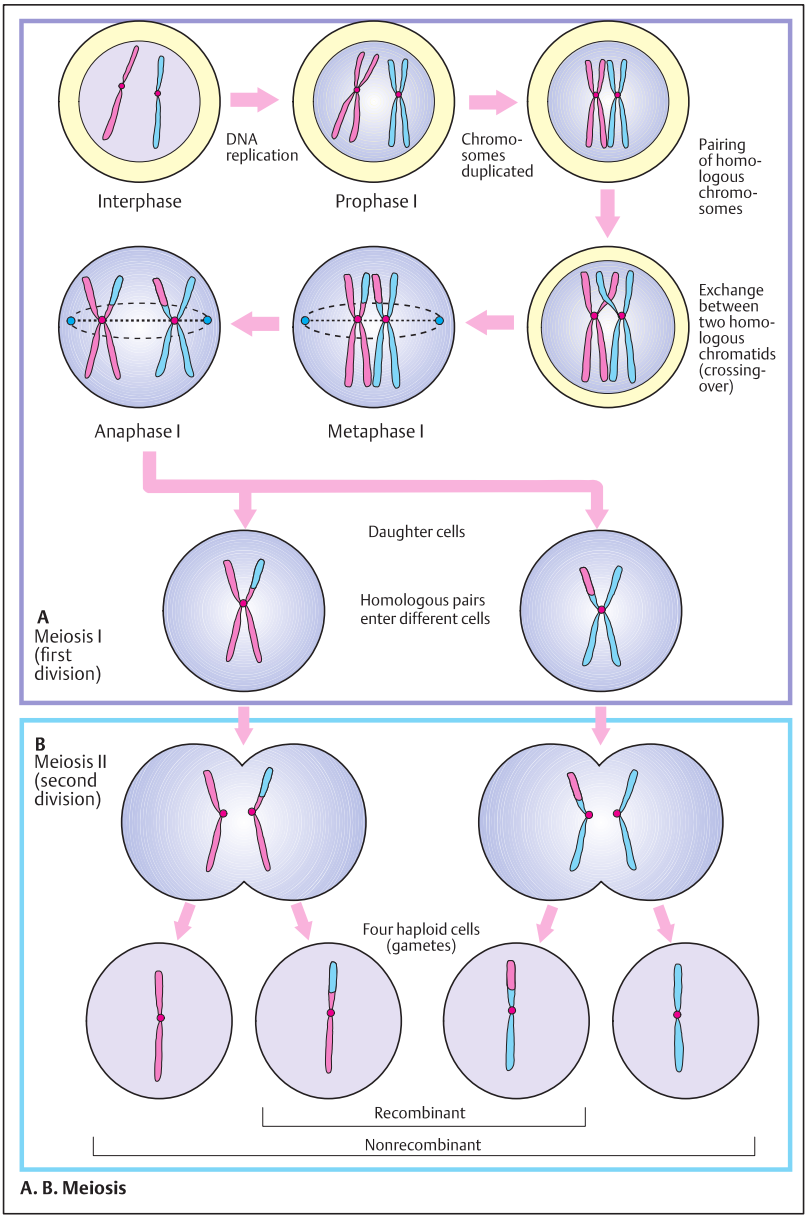
Moens PB, ed. Meiosis. New York, NY: Academic Press, 1987

Page SL, et al. Chromosome choreography: the meiotic ballet. Science 2003; 301(5634): 785–789

Petronczki M, et al. Un ménage à quatre: the molecular biology of chromosome segregation in meiosis. Cell 2003; 112(4): 423–440

Whitehouse LHK. Towards an Understanding of the Mechanism of Heredity. 3rd ed. London: Edward Arnold, 1973

Zickler D, et al. Meiotic chromosomes: integrating structure and function. Annu Rev Genet 1999; 33: 603–754


Interphase
DNA replication
Prophase I
Chromo-somes duplicated
Pairing of homo-logous chromo-somes
Exchange between two homo-logous chromatids (crossing-over)
Metaphase I
Anaphase I
Daughter cells
Homologous pairs enter different cells
A
Meiosis I (first division)
B
Meiosis II (second division)
Four haploid cells (gametes)
Recombinant
Nonrecombinant

A. B. Meiosis

Meiosis Prophase I

The decisive cytological and genetic events take place in the prophase of meiosis I. In pro-phase I, exchanges between homologous chromosomes occur regularly by crossing-over. Crossing-over, a term introduced by Morgan and Cattell in 1912, is an elaborate cytological process by which parts of chromosomes of maternal and paternal origin exchange stretches of DNA. This results in new combinations of chromosome segments (genetic recombination).

A. Prophase of meiosis I

The prophase of meiosis I passes through five consecutive stages. The first is the *leptotene* stage. Here the chromosomes first become visible as fine threadlike structures (only one chromosome pair is shown). Next is *zygotene*: each chromosome is visible as a paired structure, the result of DNA replication prior to the beginning of prophase. Consequently, each chromosome has been duplicated and consists of two identical chromatids (sister chromatids). These are held together at the centromere. Each chromatid contains a DNA double helix. Two homologous chromosomes that have paired are referred to as a bivalent. In the *pachytene* stage, the bivalents become thicker and shorter. In the *diplotene* stage, the two homologous chromosomes separate but remain attached to each other at a few points, each called a *chiasma* (see next page). In the next phase, *diakinesis*, each of the chromosome pairs has separated further, although they still remain attached to each other at the ends. A chiasma corresponds to a region at which crossing-over has taken place previously. However, in late diakinesis, the chiasmata shift distally, called *chiasma terminalization*. The mechanisms of meiosis II correspond to those of mitosis. The different stages cannot be sharply distinguished.

B. Synaptonemal complex

The synaptonemal complex, independently observed in spermatocytes by D. Fawcett and M. J. Moses in 1956, is a complex structure formed during meiotic prophase I. It consists of two chromatids (1 and 2) of maternal origin (mat) and two chromatids (3 and 4) of paternal origin (pat). It initiates chiasma formation and is the prerequisite for crossing-over and subsequent recombination. Double-strand breaks in homologous chromosomes occur prior to formation of the synaptonemal complex. (Figure adapted from Alberts et al., 2015.)

C. Chiasma formation

Chiasma is the term introduced by F. A. Janssens in 1909 for the cytological manifestation of crossing-over during meiotic prophase I. A chiasma forms between one chromatid of a chromosome of maternal origin (chromatids 1 and 2 in the figure) and one chromatid of a chromosome of paternal origin (chromatids 3 and 4). Either of the two chromatids of one chromosome can cross over with one of the chromatids of the homologous chromosome (1 and 3, 2 and 4, etc.).

D. Genetic recombination

Through crossing-over, new combinations of chromosome segments arise (recombination). As a result, recombinant and nonrecombinant chromosome segments can be differentiated. In the figure, the areas A–E (shown in pink) of one chromosome and the corresponding areas a–e (shown in blue) of the homologous chromosome become a–b–C–D–E and A–B–c–d–e, respectively.

E. Pachytene and diakinesis under the microscope

Diakinesis under the light microscope (**a**) and pachytene under the electron microscope (**b**) are shown. An extra chromosome 21 (red arrow in **a**) present in a man with trisomy with an unpaired 21. The X and Y chromosomes are paired end-to-end (see p. 230). The thickened (duplicated) chromosomes and the XY bivalent in **b** are visible. (Photographs provided by Dr. R. Johannisson, Lübeck. a from Johsnnisson et al, 1983).

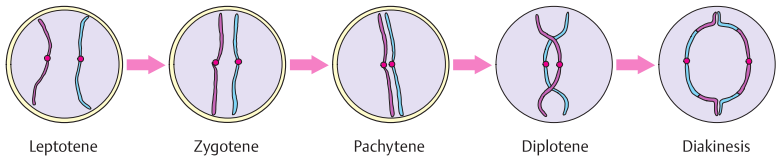
Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

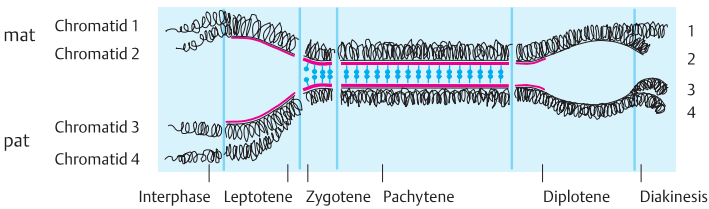
Johannisson R, et al. Down's syndrome in the male. Reproductive pathology and meiotic studies. Hum Genet 1983; 63(2): 132–138

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

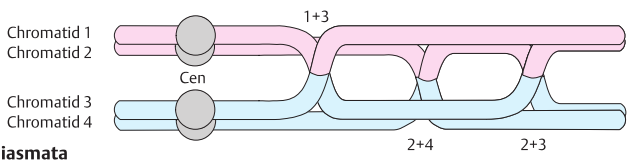
Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001



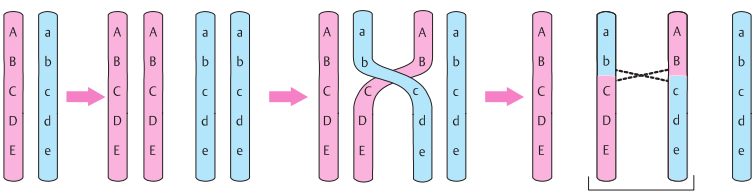
A. Prophase of meiosis I



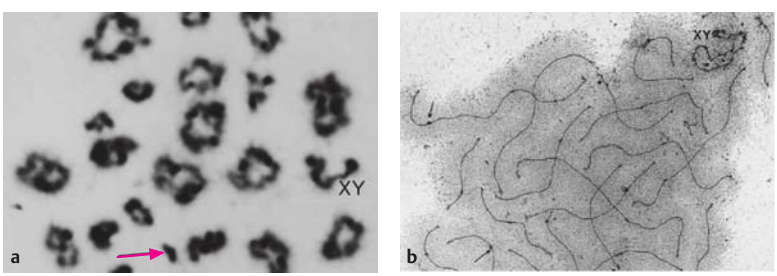
B. Synaptonemal complex



C. Chiasmata



D. Genetic recombination by crossing-over



E. Diakinesis (a, light microscope) and early pachytene (b, electron microscope)

Formation of Gametes

Gametes (germ cells) are the cells that transmit genetic information from one generation to the next. They are produced in the gonads. In females, the process is called *oogenesis* (formation of oocytes), and in males, it is called *spermatogenesis* (formation of spermatozoa). Primordial germ cells migrate during early embryonic development from the genital ridge to the gonads. Here they increase in number by mitotic divisions. The actual formation of germ cells (gametogenesis) begins with meiosis. Gametogenesis in males and females differs in duration and results.

A. Spermatogenesis

Following migration into the embryonic testis, germ cells develop into diploid spermatogonia and go through mitotic divisions. Primary spermatocytes result from the first meiotic division, beginning at the onset of puberty. At the completion of meiosis I, one primary spermatocyte gives rise to two secondary spermatocytes. Each has a haploid set of duplicated chromosomes. In meiosis II, each secondary spermatocyte divides to form two spermatids. Thus, one primary spermatocyte forms four spermatids, each with a haploid chromosome complement. The spermatids differentiate into mature spermatozoa in approximately 6 weeks. Male spermatogenesis is a continuous process. In human males, the time required for a spermatogonium to develop into a sperm cell is approximately 90 days.

B. Oogenesis

During oogenesis in early embryogenesis, germ cells migrate from the genital ridge to the ovary form oogonia by repeated mitoses. A primary oocyte results from the first meiotic division of an oogonium. In human females, meiosis I begins approximately 4 weeks before birth. Then meiosis I is arrested in a stage of prophase designated *dictyotene*. The primary oocyte persists in this stage until ovulation when meiosis I is continued. In primary oocytes, the cytoplasm divides asymmetrically in both meiosis I and meiosis II, which results in two cells of unequal sizes. The larger will form the egg; the smaller cell becomes a polar body, not a germ cell. When a secondary oocyte divides, one daughter cell becomes an oocyte and the other a polar body II. The polar bodies degenerate and do not develop. In the secondary oocyte, each chromosome still consists of two sister chromatids, which separate during meiosis II. At ovulation, the secondary oocyte is released from the ovary, and if fertilization occurs, meiosis is then completed.

Medical relevance

Faulty distribution of the chromosomes (nondisjunction) during meiosis I or meiosis II is the cause of aberrations of the chromosome number (see p. 382).

A polar body may be examined to assess whether a genetic abnormality is present in the fetus. On rare occasions, a polar body may become fertilized. This can give rise to an incompletely developed twin.

The mutation rate differs between oogenesis and spermatogenesis. The difference in time in the formation of gametes during oogenesis and spermatogenesis is reflected in the difference in germline cell divisions. The number of cell divisions in spermatogenesis and oogenesis differs considerably. On average, approximately 380 chromosome replications take place in the progenitor cells of spermatozoa by the age of 30 years, and approximately 610 chromosome replications by the age of 40 years. Altogether, 25 times more cell divisions occur during spermatogenesis than during oogenesis (Crow, 2000). This probably accounts for the higher mutation rate in males, especially with increased paternal age. In the female, an average number of 22 mitotic cell divisions occurs before meiosis, resulting in a total of 23 chromosome replications.

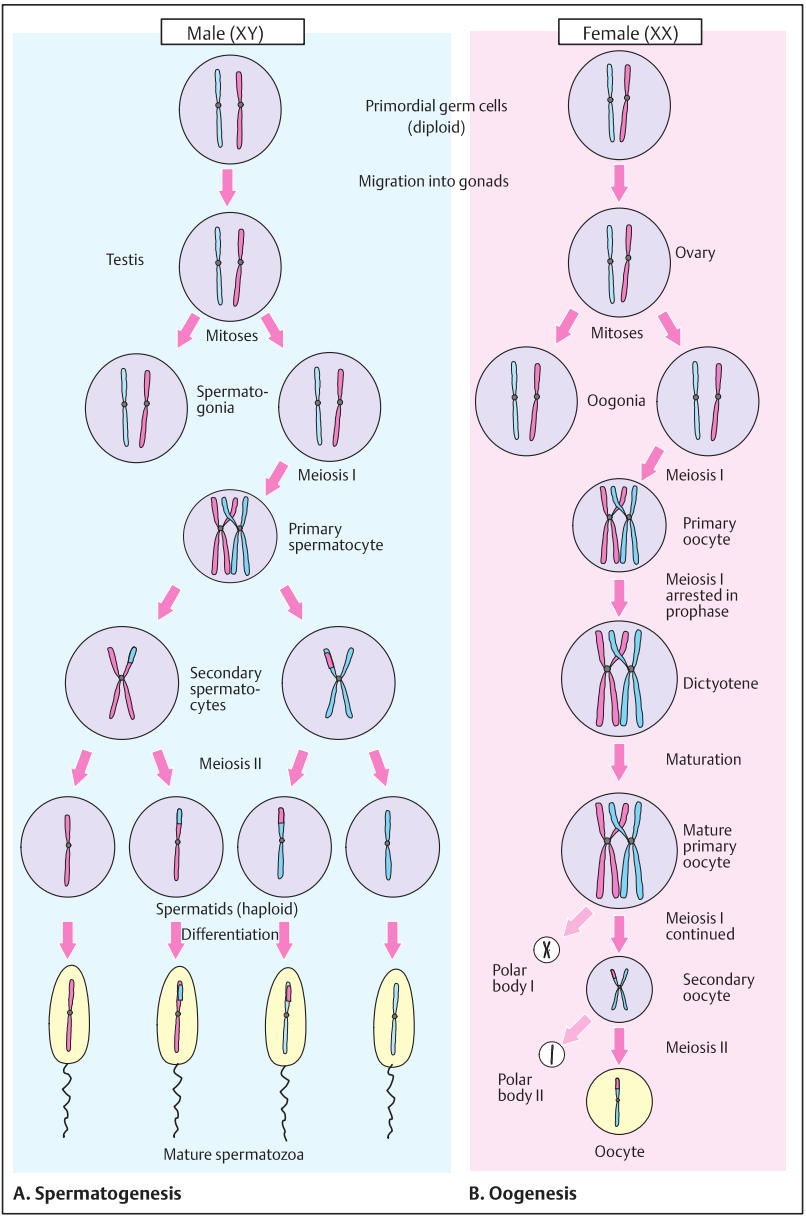
From a total number of germ cells in the ovary of the human fetus at approximately the fifth month of 6.8×10^6, this has been reduced to 2×10^6 by the time of birth and to approximately 200,000 by puberty. Of these, approximately 400 eventually go through ovulation.

Further Reading

Crow JF. The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet 2000; 1(1): 40–47

Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001

Nachman MW, et al. Estimate of the mutation rate per nucleotide in humans. Genetics 2000; 156(1): 297–304



A. Spermatogenesis **B. Oogenesis**

Programmed Cell Death

At certain stages in the development of a multicellular organism, some cells must die instead of developing into differentiated cells. This process is called *apoptosis* (programmed cell death, derived from the Greek word *apo*, meaning from/off/without, and *ptosis*, meaning fall), as suggested by Kerr in 1972. The importance of this biological phenomenon was first realized in studies of a tiny worm, the soil nematode *Caenorhabditis elegans* (see p. 202). Apoptosis is regulated in several apoptosis pathways by many proteins, which either induce or prevent apoptosis. Apoptosis can be triggered from outside the cell (extrinsic pathway) or from within the cell (intrinsic pathway).

A. Importance of apoptosis

Apoptosis occurs mainly during embryonic development. For example, the digits in the developing mammalian embryo are sculptured by apoptosis (**1**). The paws (hands) start out as spade-like structures. The formation of digits requires that cells between them die (here shown as bright green dots on the left). More staggering is the amount of apoptosis in the developing vertebrate nervous system. Normally, up to half of the nerve cells die soon after they have been formed. In the embryos of mice that lack an important gene regulating apoptosis (the gene encoding caspase 9, see below), neurons proliferate excessively and the brain protrudes above the face (**2**). (Images from Alberts et al, 2015, and Gilbert & Barresi, 2016).

B. Cellular events in apoptosis

The first visible signs of apoptosis are condensation of chromatin and shrinking of the cell. The cell membrane shrivels (membrane blebbing), and the cell begins to disintegrate (nuclear segmentation, DNA fragmentation). Apoptotic bodies of cell remnants form and are eliminated by phagocytosis.

C. Regulation of apoptosis

Specialized cysteine-containing aspartate proteinases, called *caspases*, activate or inactivate each other in a defined sequence. Binding of a ligand, Fas, of a cytotoxic T cell (see section on immune system) to the Fas receptor (also called *CD95*) activates the intracellular adaptor protein FADD (Fas-associated death domain). This binds to and activates procaspase 8 into active caspase 8. Caspase 8 causes release of cytochrome c in mitochondria (see p. 232) and activates several different effector caspases.

The mouse and human genomes contain 13 caspase genes (1–12). Human caspases 3 and 6 to 10 are involved in apoptosis; the others are involved in inflammation. Caspase 8 also serves as a selective signal transducer for nuclear factor kappa B during the early genetic response to an antigen. Other regulators of apoptosis are members of the Bcl-2 family.

Medical relevance

Mutations in genes encoding inhibitors of apoptosis or apoptosis-inducing factors cause several different types of tumors (see "apoptosis" in OMIM), for example, B-cell lymphoma (OMIM 151430) caused by a defective bcl protein due to a mutation in the *BCL2* gene.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Danial NN, Korsmyer SJ. Cell death: critical control points. Cell 2004; 116(2): 205–219

Gilbert SF, Barresi MJF. Developmental Biology. 11th ed. Sunderland, MA: Sinauer, 2016

Hengartner MO. The biochemistry of apoptosis. Nature 2000; 407(6805): 770–776

Hotchkiss RS, et al. Cell death. N Engl J Med 2009; 361(16): 1570–1583

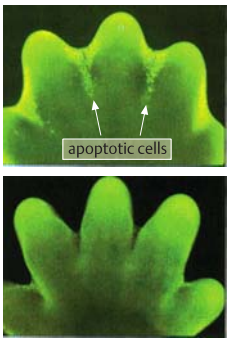
Kerr JF, et al. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer 1972; 26(4): 239–257

Koolman J, Roehm KH. Color Atlas of Biochemistry. 2nd ed. Stuttgart: Thieme, 2005

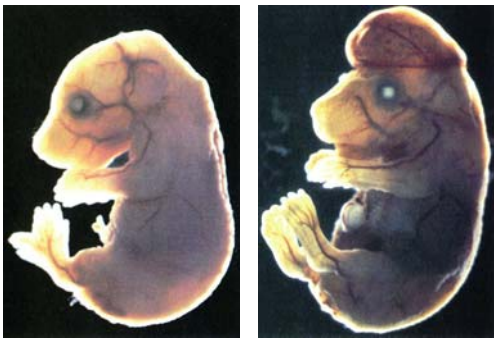
Kuida K, et al. Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. Cell 1998; 94(3): 325–337

Nagata S. DNA degradation in development and programmed cell death. Annu Rev Immunol 2005; 23: 853–875

Su H, et al. Requirement for caspase-8 in NF-kappaB activation by antigen receptor. Science 2005; 307 (5714): 1465–1468

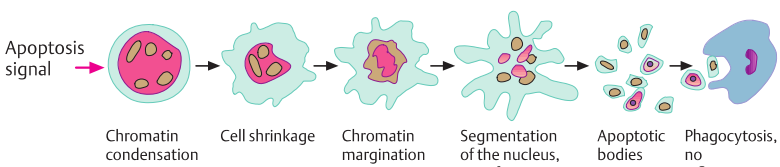


1. Apoptosis in the paw of a mouse embryo

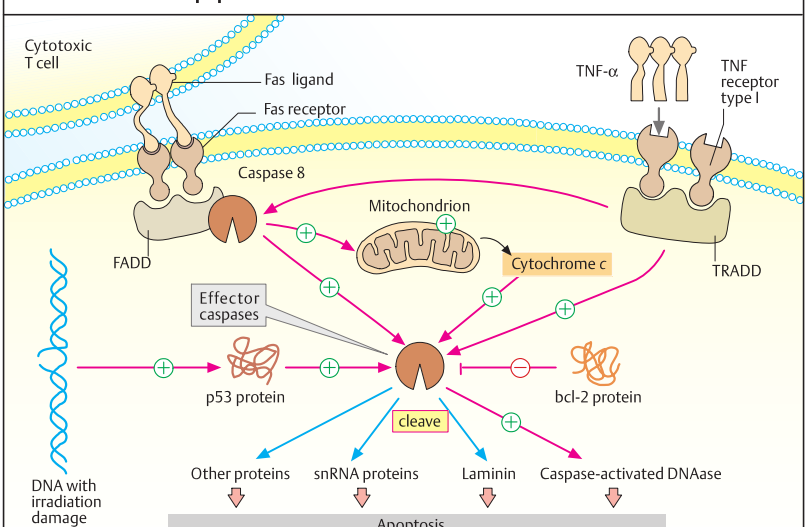


2. Disruption of brain development (right) in a mouse embryo lacking caspase 9

A. Importance of apoptosis



B. Cellular events in apoptosis



C. Regulation of apoptosis

Cultured Cells

Cells of animals and plants can live and multiply outside the body in a cell culture. This usually represents an undifferentiated cell type, fibroblasts. These can be grown into colonies of identical cells (cell cloning) as a source of homogeneous cells. This allows experimental manipulations that are not feasible in living organisms. Use of cell cultures began in 1940 when W. Earle established a permanent mouse cell strain, and T. T. Puck and coworkers grew clones of human cells in vitro. In 1965, this led to the development of *somatic cell genetics*.
Cultured cells in vitro have a finite life span as cells in vivo, as outlined in Genetic Background of Aging Processes earlier (see p 34).

A. Skin fibroblast culture

A typical cell culture requires a dish with medium at 37°C containing vitamins, sugar, serum (containing numerous growth factors and hormones), the nine essential amino acids for vertebrates, and usually glutamine and cysteine. The predominant cell type that grows from a piece of mammalian tissue in culture is the fibroblast. To initiate a culture, a small piece of skin (2 × 4 mm) is obtained in sterile conditions and cut into smaller pieces, which are placed in a culture dish where they attach to the bottom (adhesion culture). After about 8-14 days, cells grow out and form a monolayer. They stop dividing because of contact inhibition (this is lost in tumor cells). When transferred into new culture vessels (subcultures), the cells will resume growing until they again become confluent. With a series of subcultures, several million cells can be obtained for a given study. Cultured cells are highly sensitive to increased temperature and do not survive above approximately 39°C, whereas in special conditions, they can be stored alive in vials kept in liquid nitrogen at –196°C. They can be thawed after many years, or even decades, and cultured again.

B. Hybrid cells

Cells in culture can be induced to fuse by exposure to polyethylene glycol or Sendai virus. If parental cells from different species are fused, interspecific (from different animal species) hybrid cells can be derived. The hybrid cells can be distinguished from the parental cells by using parental cells deficient in thymidine kinase (TK^-) or hypoxanthine phosphoribosyltransferase ($HPRT^-$). When cell cultures of parental type A (TK^-, **1**) and type B ($HPRT^-$, **2**) are cultured together (cocultivation, **3**), some cells fuse (**4**). In a selective medium containing hypoxanthine, aminopterin, and thymidine (HAT medium; Littlefield, 1964), only fused cells with a nucleus from each of the parental cells (**1** and **2**) can grow. Cells that have not fused cannot grow in HAT medium (**5**). The reason for this is that TK^- cells cannot synthesize thymidine monophosphate; $HPRT^-$ cells cannot synthesize purine nucleoside monophosphates. The fused cells complement each other. The two nuclei of fused cells (heterokaryon) will also fuse (**6**). This forms a hybrid cell (**7**). These can be propatated in a culture for further study (**8**). Studies of hybrid cells have yielded genetic insights into cell functions, which would not have been available otherwise.

C. Radiation hybrids

Human chromosome fragments (**1**) obtained from irradiated cells with lethal roentgen doses of 3 to 8 Gy were fused with rodent cells containing intact chromosomes (**2**). Some human chromosome fragments will be integrated into the rodent chromosomes (**3**). Those cells that contain human DNA can be identified by human chromosome-specific probes after culture in a selective medium. A random selection of human DNA fragments contained in a radiation hybrid panel was used to construct long-range maps of chromosomes.

Further Reading

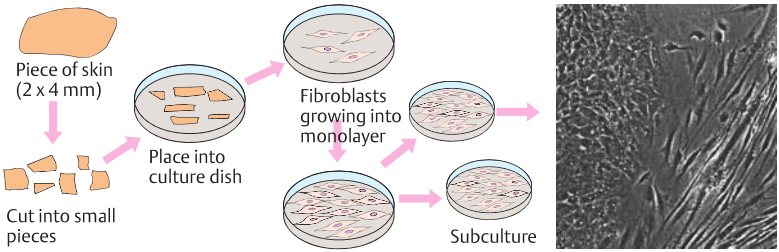
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Hayflick L. Mortality and immortality at the cellular level. A review. Biochemistry (Mosc) 1997; 62(11): 1180–1190

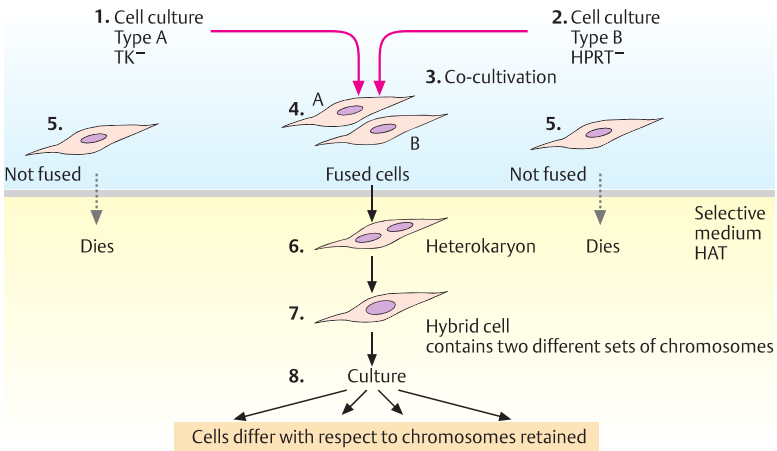
Littlefield JW Selection of hybdrids from matings of fibroblasts in vitro and their presumed recombinants. Science 1964; 145(3633): 709–710

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

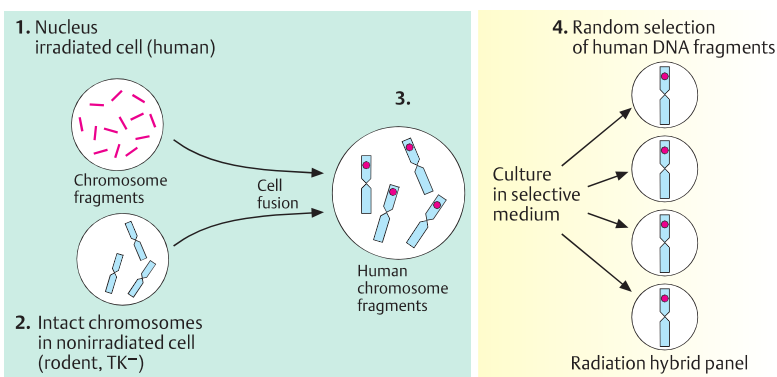
McCarthy LC. Whole genome radiation hybrid mapping. Trends Genet 1996; 12(12): 491–493



A. Skin fibroblast culture



B. Hybrid cells for study



C. Radiation hybrids

The Mendelian Traits

The scientific foundation for the rules underlying inheritance was laid in 1865 by the Augustinian monk Gregor Mendel. He presented remarkable observations at the Naturgeschichtliche Vereinigung von Brünn (The Natural History Society of Brünn/Brno), published in 1866. In his work, entitled *Versuche über Pflanzenhybriden* (Experiments with Hybrid Plants), Mendel observed that certain traits in the garden pea (*Pisum sativum*) are inherited according to regular patterns and independently of one another. It was not until 1900 that H. de Vries, C. Correns, and E. Tschermak independently recognized the importance of Mendel's discovery for biology.

A. The pea plant (*Pisum sativum*)

The garden pea normally reproduces by self-fertilization. Pollen from the anther falls onto the stigma of the same blossom. However, one can easily cross-fertilize (cross-pollinate) pea plants. The plant (left) consists of a stem, leaves, blossoms, and seedpods. In the blossom (right), the female and male reproductive organs are visible. The female pistil comprises stigma, style, and ovule. The male organ is the stamen, comprising the anther and filament. For cross-fertilization, Mendel opened a blossom and removed the anther to avoid self-fertilization. Then he transferred pollen from another plant to the receptive stigma directly.

B. The traits observed by Mendel

Mendel observed seven characteristic traits: height of the plants (**1a**), location of the blossoms on the stem of the plant (**2a**), the color of the pods (**3**), the form of the pods (**4**), the form of the seeds (**5**), the color of the seeds (**6**), and the color of the seed coat (**7**). These traits were transmitted in defined proportions to the next plant generation.

Deviations from the Mendelian pattern of inheritance

Mendelian traits can deviate from the proportions described on the next page. *Epistasis* is a nonreciprocal interaction between nonallelic genes. As a result, the effect of one gene masks the expression of alleles of another gene. In 1902, W. Bateson described this phenomenon in the recessive gene *apterous* (*ap*) in *Drosophila.* Homozygotes are wingless, but, in addition, other genes affecting wing morphology, such as curled wing, are masked (the *ap* gene is epistatic to curled wing). The Bombay blood group is another example (Bhende et al., 1952; Race and Sanger, 1975).

Meiotic drive refers to the preferential transmission of one allele over others. As a result, one trait occurs in offspring much more frequently than others. Striking examples are the t complex of the mouse (~99% instead of 50% of offspring of heterozygous t/+ male mice are also heterozygous) and segregation disorder in *Drosophila.*

Genomic imprinting (see p. 182) is a further cause of deviation from the Mendelian pattern of inheritance.

Medical relevance

Several thousands of human diseases or their disposition for occurrence are transmitted by the Mendelian rules of inheritance described on the following pages.

Further Reading

Bhende YM, et al. A "new" blood group character related to the ABO system. Lancet 1952; 1(6714): 903–904

Brink RA, Styles ED. Heritage from Mendel. Madison, WI: University of Wisconsin Press, 1967

Corcos AF, Monaghan FV. Gregor Mendel's Experiments on Plant Hybrids. New Brunswick, NJ: Rutgers University Press, 1993

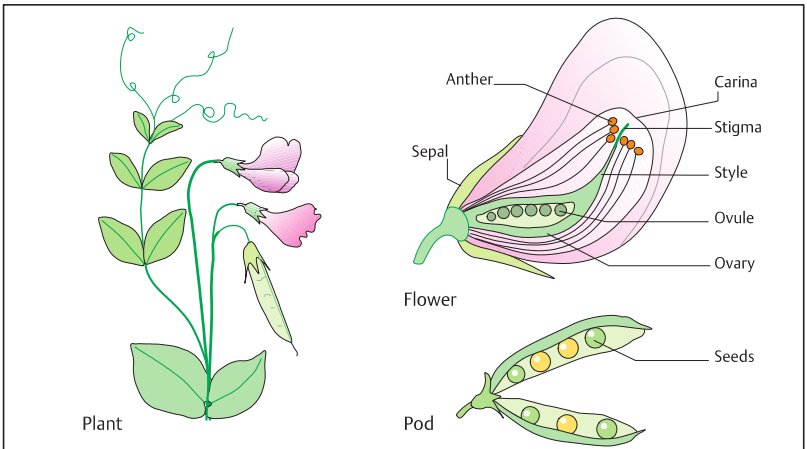
Griffith AJF, et al. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2015

Mendel G. Versuche über Pflanzenhybriden. Verh naturf Ver Brünn 1866; 4: 3–47

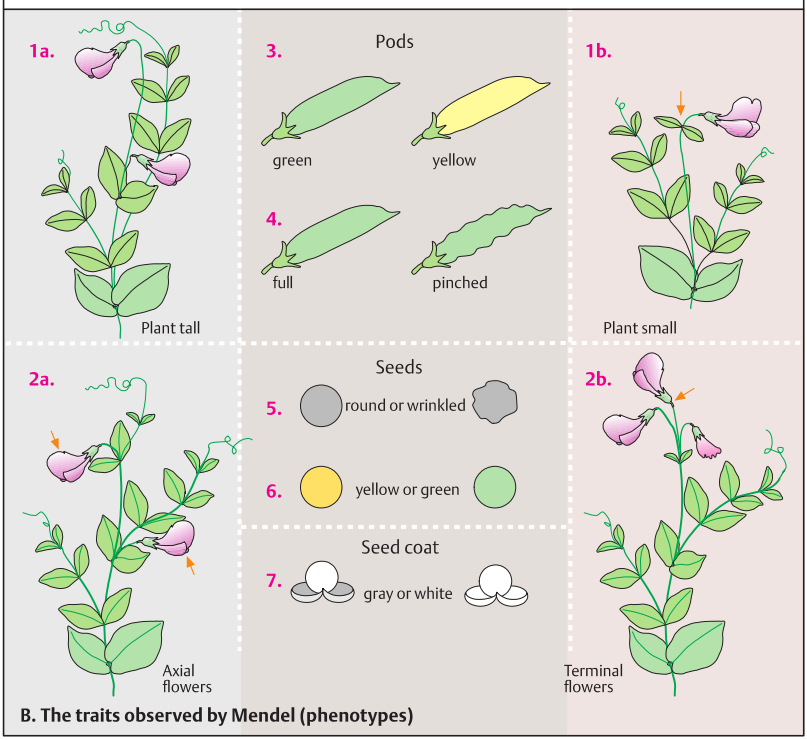
Race RR, Sanger R. Blood Groups in Man. 6th ed. Oxford: Blackwell, 1975

Vogel F, Motulsky AG. Human Genetics. Problems and Approaches. 3rd ed. Heidelberg: Springer-Verlag, 1997

Weiling F. Johann Gregor Mendel: Der Mensch und Forscher. II Teil. Der Ablauf der Pisum Versuche nach der Darstellung. Med Genetik 1993; 2: 208–222



A. The pea plant (*Pisum sativum*)



B. The traits observed by Mendel (phenotypes)

Transmission to the Next Generation

Mendel observed that the traits of the pea plant (*Pisum sativum*) (described on the previous page) were transmitted to the next generation according to a defined pattern. They reappeared in the next generation in certain ratios (genetic segregation). He provided a meaningful biological interpretation, later called *Mendel's laws of inheritance*.

A. Segregation of dominant and recessive traits

Mendel crossed plants of the parental generation P with seeds with different shapes (smooth or wrinkled) and colors (yellow or green). He observed that in the first filial (daughter) generation, F_1, all seeds were smooth and yellow. In the next generation, F_2, which arose by self-fertilization, the traits observed in the P generation (smooth, wrinkled, green and yellow) reappeared. Among 7,324 seeds of one experiment, 5,474 were smooth and 1,850 were wrinkled. In the seeds with different colors (green versus yellow), in a total of 8,023 seeds of the F_2 generation, 6,022 were yellow and 2,001 green. In both cases, this corresponded to a ratio of 3:1.

The only traits, round and yellow, which Mendel observed in the F_1 generation, he called dominant. The traits that did not appear in the F_1 generation (wrinkled or green) he called *recessive*. His observation that a dominant and a recessive pair of traits occurring (segregating) in the F_2 generation in the ratio 3:1 is known as the *first law of Mendel*.

B. Backcross of an F_1 hybrid with a parent plant

When Mendel backcrossed the F_1 plant round with a parent plant wrinkled, the recessive trait (**1**), both traits occurred in the next generation in a ratio of 1:1 (106 round and 102 wrinkled). This is called the *second law of Mendel*.

The interpretation of this experiment (**2**) was that the F_1 plant (round) contains two traits: one for round (R, dominant over wrinkled, r) and one for wrinkled (r, recessive to round, R). This plant is *heterozygous* (Rr) and therefore can form two types of gametes (R and r).

In contrast, the other plant is *homozygous* (rr) for wrinkled (r). It can form only one type of gamete (r, wrinkled). Half of the offspring of the heterozygous plant receive the dominant trait (R, round), and the other half the recessive trait (r, wrinkled). The resulting traits R and r were distributed in a ratio of 1:1, or 50% each. The observed trait is called the *phenotype* (the observed appearance of a particular characteristic). The composition of the two factors (genes) R and r, (Rr) or (rr), is called the *genotype*. The alternative forms of a trait (here, round and wrinkled) are called *alleles*. They are the result of different genetic information at one given gene locus.

If the alleles are different, the genotype is heterozygous; if they are the same, it is homozygous (this statement is always in reference to a single, given gene locus).

The numbers representing the ratios given in part A have been subject of a controversy raised by the famous statistician and geneticist Sir Ronald Fisher. He suggested in 1936 that Mendel's experimental results were too close to be expected when considering statistical fluctuations. Fisher raised the question whether Mendel might have introduced a bias in favor of the ratios he observed, at least unconsciously. Pires and Branco provided a statistical model that could explain the situation without assuming that Mendel changed his data in an unacceptable way.

Further Reading

Brink RA, Styles ED. Heritage from Mendel. Madison, WI: University of Wisconsin Press, 1967

Corcos AF, Monaghan FV. Gregor Mendel's Experiments on Plant Hybrids. New Brunswick, NJ: Rutgers University Press, 1993

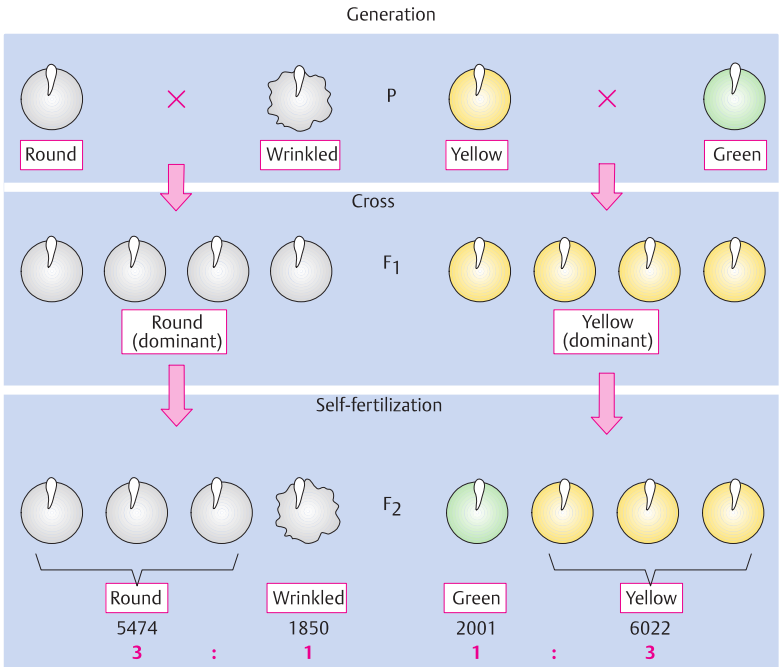
Fisher RA. Has Mendel's work been rediscovered? Ann Sci 1936; 1(2): 115–137

Griffith AJF, Wessler SR, Carroll SB, Doebley JF. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2015

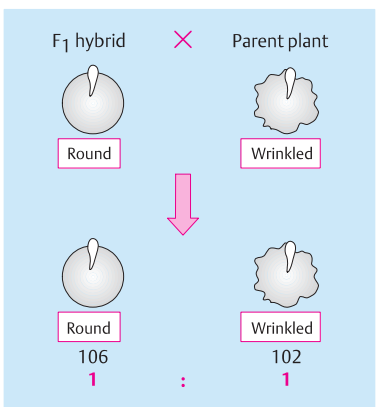
Mendel G. Versuche über Pflanzenhybriden. Verh naturf Ver Brünn 1866; 4: 3–47

Pires AM, Branco JA. A statistical model to explain the Mendel-Fisher controversy. Stat Sci 2010; 25 (4): 545–565

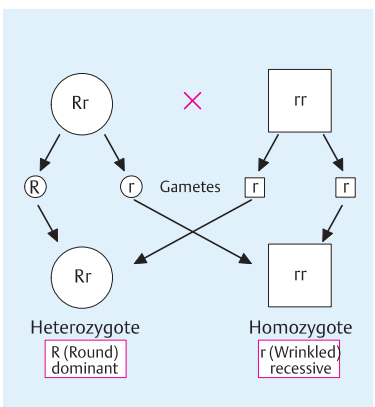
Vogel F, Motulsky AG. Human Genetics. Problems and Approaches. 3rd ed. Heidelberg: Springer-Verlag, 1997



A. Segregation of dominant and recessive traits



1. Experiment



2. Interpretation

B. Backcross of an F_1 hybrid with a parent plant

Independent Distribution

In a further experiment, Mendel observed that the two different traits (round/wrinkled and yellow/green) are inherited independently of each other. Each pair of traits showed the same 3:1 distribution of the dominant over the recessive trait in the F_2 generation as he had observed previously. The segregation of two pairs of traits again occurred in defined ratios.

A. Independent distribution of two traits

When Mendel crossed plants with round and yellow seeds with plants with wrinkled and green seeds, only round and yellow seeds showed up in the F_1 generation. Of 556 plants in the F_2 generation, the two pairs of traits occurred in the following distribution: 315 seeds yellow and round, 108 seeds yellow and wrinkled, 101 seeds green and round, and 32 seeds green and wrinkled. This corresponds to a segregation ratio of 9:3:3:1. This is referred to as the *third Mendelian law*.

B. Interpretation of the observations

Mendel's observations can be summarized as follows. If we assign the capital letter G to the dominant gene *yellow*, a lowercase g to the recessive gene *green*, the capital letter R to the dominant gene *round*, and a lowercase r to the recessive gene *wrinkled*, the following nine genotypes of these two traits can occur: **GGRR, GGRr, GgRR, GgRr** (all *yellow* and *round*); **GGrr, Ggrr** (*yellow* and *wrinkled*); **ggRR, ggRr** (*green* and *round*); and **ggrr** (*green* and *wrinkled*). The distribution of the traits shown in A is the result of the formation of gametes of different types, that is, depending on which of the genes they contain.

The ratio of the dominant trait yellow (**G**) to the recessive trait green (**g**) is 12:4, that is, 3:1. Also, the ratio of dominant round (**R**) to wrinkled (**r**) seeds is 12:4, that is, 3:1.

The results can be visualized in a diagram called the *Punnett square*. This is a checkerboard way of determining the types of zygotes produced when two gametes with a defined genotype fuse. It was first published in a book entitled Mendelism in its second edition by R.C. Punnett in 1907.

This square shows the nine different genotypes that can be formed in the zygote after fertilization. Altogether, there are 9/16 yellow round seeds (**GRGR, GRGr, GrGR, GRgR, gRGR, GRgr, GrgR, gRGr, grGR**), 3/16 green round (**gRgR, gRgr, grgR**), 3/16 yellow wrinkled (**GrGr, Grgr, grGr**), and 1/16 green wrinkled seeds (**grgr**). Each of the two traits (dominant yellow versus recessive green, or dominant round versus recessive wrinkled) occurs in a 3:1 ratio (dominant versus recessive).

Mendel's observations differed fundamentally from all other 19th-century attempts to understand heredity. First, Mendel simplified the experimental approach by selecting traits that could be easily observed; second, he assessed the pattern of transmission from one generation to the next quantitatively; and, third, he provided a biologically meaningful interpretation by noting that each pair of traits was inherited independently of the other pairs of traits in predictable pattern. This fundamentally new insight into the process of heredity and its significance was not recognized until 1900. Then three scientists, Hugo de Vries (Amsterdam), Erich Tschermak (Vienna) and Carl Correns (Tübingen), independently recognized the fundamental nature of Mendel's work.

Today we know that genetically determined traits are independently inherited (segregation) only when they are located on different chromosomes or are far enough apart on the same chromosome to be separated each time by recombination. Otherwise they might stay together (see linkage, p. 116).

Further Reading

Brink RA, Styles ED. Heritage from Mendel. Madison, WI: University of Wisconsin Press, 1967

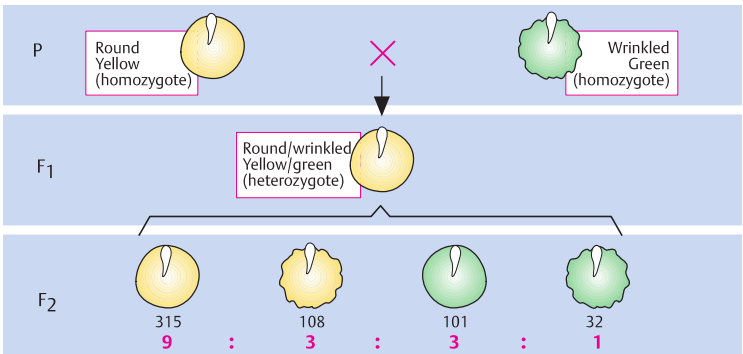
Edwards AW. Punnett's square. Stud Hist Philos Biol Biomed Sci 2012; 43: 219–224

Griffith AJF, Wessler SR, Carroll SB, Doebley JF. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2015

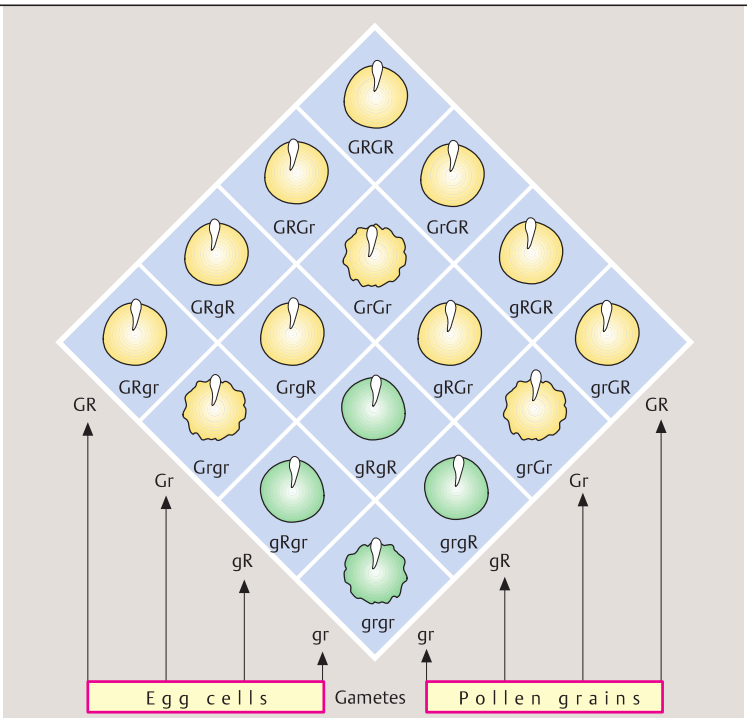
Mendel G. Versuche über Pflanzenhybriden. Verh naturf Ver Brünn 1866; 4: 3–47

Punnett RC. Mendelism. 2nd ed. Cambridge, MA: Bowes & Bowes, 1907

Vogel F, Motulsky AG. Human Genetics. Problems and Approaches. 3rd ed. Heidelberg: Springer-Verlag, 1997



A. Independent segregation of two traits



B. Interpretation of the observation

Phenotype and Genotype: Application in Genetic Counseling

Mendel's observations are directly applied when assessing a possible risk for the occurrence of an inherited human disease in a family during genetic counseling. In doing so, one examines the genetic relationship of individuals presented in a pedigree (pedigree analysis). An observed *phenotype* could be a disease, a blood group, a protein variant, a laboratory result, or any other attribute determined by observation.

A. Symbols in a pedigree drawing

The symbols shown here represent a common way of drawing a pedigree in a family. Usually, this should include at least three generations. Only two are shown here. Males are shown as squares, females as circles. Individuals of unknown sex (e.g., because of inadequate information) are shown as diamonds. In medical genetics, the degree of reliability in determining the phenotype, for example, the presence or absence of a disease, should be stated. In each case, it must be stated which phenotype (e.g., which disease) is being dealt with. Established diagnoses (data complete), possible diagnoses (data incomplete), and questionable diagnoses (statements or data dubious) should be differentiated.

B. Genotype and phenotype

The definitions of genotype and phenotype refer to a given gene locus. Different forms of genetic information at one and the same gene locus are called *alleles*. In diploid organisms—all animals and many plants—there are three possible genotypes with respect to two alleles at any one locus: (i) *homozygous* for two identical alleles, (ii) *heterozygous* for the two different alleles, and (iii) *homozygous* for the other two identical alleles.

If they can be recognized in the heterozygous state, they are called *dominant*. If they can be recognized in the homozygous state only, they are *recessive*. If the two alleles can both be recognized in the heterozygous state, they are designated *codominant* (e.g., the alleles A and B of the blood group system ABO; O is recessive to A and B). The concepts dominant and recessive are attributes of the accuracy in observation and do not apply at the molecular level.

Medical relevance

The Mendelian pattern of inheritance provides the foundation for genetic counseling of patients with monogenic diseases. *Genetic counseling* is a communication process relating to all aspects related to the occurrence of a genetic disorder, in particular the diagnosis and assessment of the potential occurrence of a genetically determined disease in a family and in more distant relatives. The individual affected with a disease, who first attracted attention to a particular pedigree, is called the *index patient* (or proposita if female, propositus if male). The person who seeks information is called the *consultand*. The index patient and the consultand are very often different persons.

The goal of genetic counseling is to provide comprehensive information about the expected course of the disease, medical care, and possible treatments or an explanation, for why treatment is not possible. Genetic counseling includes a review of possible decisions about family planning as a consequence of a genetic risk. Professional confidentiality must be observed. The counselor makes no decisions. The increasing availability of information about a disease based on a DNA test (predictive DNA testing) prior to disease manifestation requires the utmost care in establishing whether it is in the interest of a given individual to have a test performed.

Further Reading

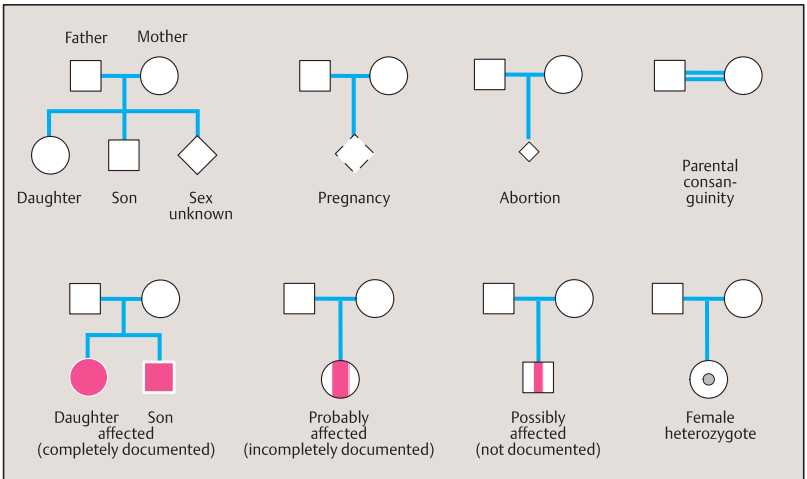
Griffith AJF, Wessler SR, Carroll SB, Doebley JF. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2015

Harper PS. Practical Genetic Counselling. 7th ed. London: Edward Arnold, 2010

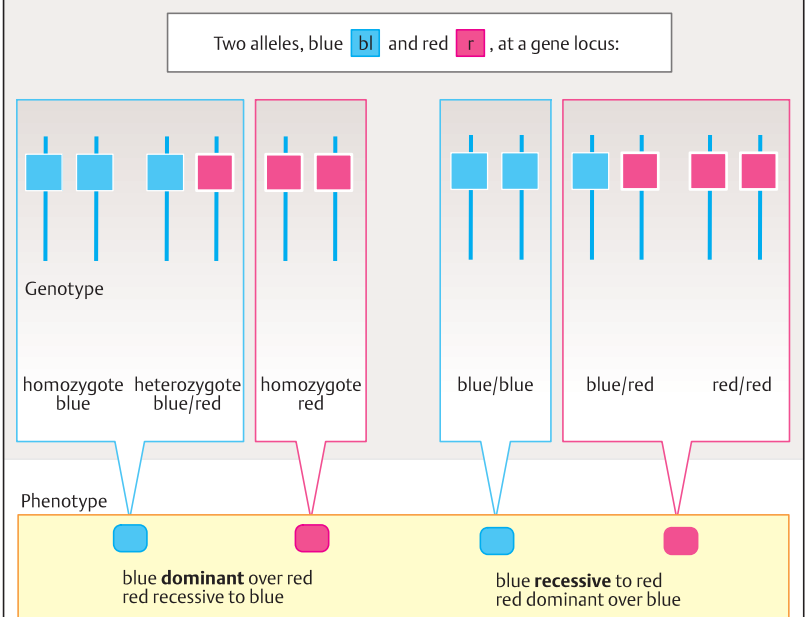
Jameson JL, Kopp P. Principles of human genetics. In: Longo DL, et al, eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 486–509

Rimoin DL, Pyeritz R, Korf B, eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. Philadelphia, PA: Elsevier-Churchill Livingstone, 2013 (e-book)

Speicher M, Antonarakis SE, Motulsky AG, eds. Vogel and Motulsky's Human Genetics. Problems and Approaches. 4th ed. Heidelberg: Springer-Verlag, 2010



A. Symbols used in a pedigree



B. Genotype and phenotype

Segregation of Parental Genotypes

The segregation (distribution) of the genotypes of the parents (parental genotypes) to their offspring depends on the combination of the alleles present in the parents. Depending on the effect of the genotype on the phenotype in the heterozygous state, an allele is classified as dominant or recessive. Hence, there are three basic modes of inheritance: (i) autosomal dominant, (ii) autosomal recessive, and (iii) X-chromosomal. For genes on the X chromosome, it is usually not important to distinguish dominant and recessive (see next page). Since the Y chromosome bears few disease-causing genes, Y-chromosomal inheritance can usually be ignored with regard to monogenic diseases.

A. Possible mating types of genotypes

For a gene locus with two alleles, there are six possible combinations of parental genotypes (**1–6**). Here two alleles, blue (bl) and red (r), are shown, with blue dominant over red. In three of the parental combinations (**1, 3, 4**), neither of the parents is homozygous for the recessive allele (red). In three parental combinations (**2, 5, 6**), one or both parents manifest the recessive allele because they are homozygous. The distribution patterns of genotypes and phenotypes in the offspring of the parents are shown in **B**. In these examples, the sex of the parents is interchangeable.

B. Expected Distribution in the offspring of parents with two alleles: A and a

With three of the parental mating types for the two alleles **A** (dominant over **a**) and **a** (recessive to **A**), there are three combinations that lead to segregation (separation during meiosis) of allelic genes. These correspond to the parental combinations 1, 2, and 3 shown in **A**. In mating types 1 and 2, one of the parents is a heterozygote (**Aa**) and the other parent is a homozygote (**aa**). The distribution of observed genotypes expected in the offspring is 1:1; in other words, 50% (0.50) are **Aa** heterozygotes and 50% (0.50) are **aa** homozygotes.

If both parents are heterozygous **Aa** (mating type 3 in A), the proportions of expected genotypes of the offspring (**AA, Aa, aa**) occur in a ratio of 1:2:1 (25%:50%:25%).

C. Distribution of genotypes

One dominant allele (in the first pedigree, **A**, in the father) can be expected in 50% of the offspring. If both parents are heterozygous, 25% of the offspring will be homozygous **aa**. If both parents are homozygous—one for the dominant allele **A**, the other for the recessive allele **a**—then all offspring are obligate heterozygotes (i.e., must necessarily be heterozygotes).

Medical relevance

An attempt to determine the mode of inheritance of a disorder (pedigree analysis) is an important fundamental approach in genetic diagnostics and counseling.

Further Reading

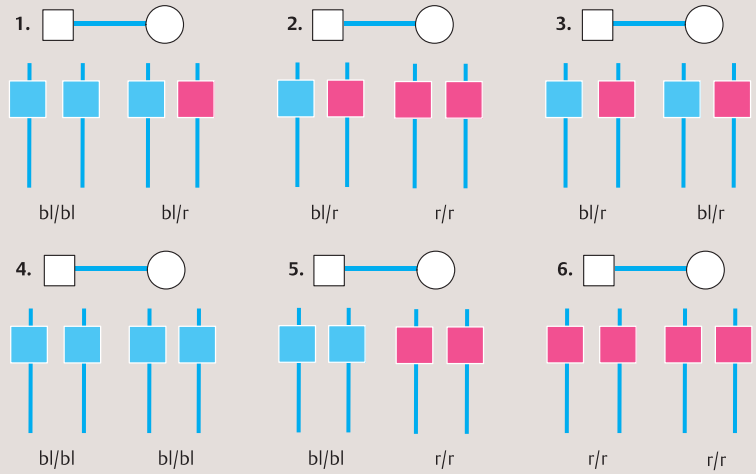
Griffith AJF, Wessler SR, Carroll SB, Doebley JF. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2007

Harper PS. Practical Genetic Counselling. 7th ed. London: Edward Arnold, 2010

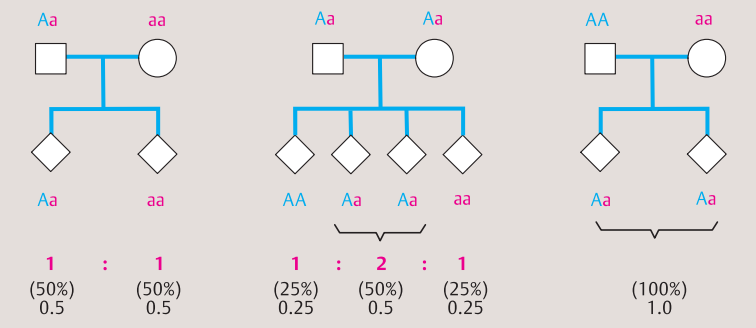
Speicher M, Antonarakis SE, Motulsky AG, eds. Vogel and Motulsky's Human Genetics. Problems and Approaches. 4th ed. Heidelberg: Springer-Verlag, 2010

Expected distribution of genotypes in the offspring of different parental genotype combinations

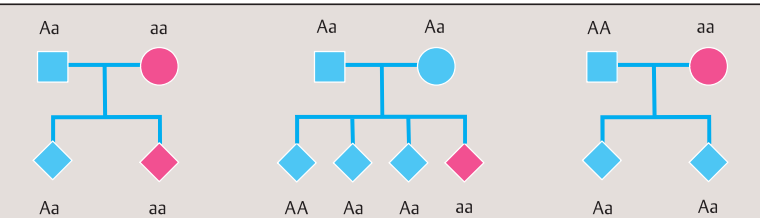
Parents	Offspring	Expected genotype proportion
AA × AA	AA	1 (100%)
AA × Aa	AA, Aa	1:1 (each 50%)
Aa × Aa	AA, Aa, aa	1:2:1 (25, 50, 25%)
AA × aa	Aa	1 (100%)
Aa × aa	Aa, aa	1:1 (50% each)
aa × aa	aa	1 (100%)



A. Possible mating types of genotypes for two alleles (blue dominant over red)



B. Expected distribution of genotypes in offspring of parents with two alleles, A and a



C. Phenotypes and genotypes in the offspring of parents with a dominant allele A and a recessive allele a

Monogenic Inheritance

Four types of monogenic inheritance can be distinguished: autosomal dominant, autosomal recessive, X-chromosomal, and Y-chromosomal. Since the human Y chromosome contains no general disease-causing genes, it can usually be ignored. In human pedigree analysis, consecutive generations are designated by roman numerals, starting with I for the first generation about which some information is available, and Arabic numerals for each individual within a generation, or nonoverlapping numbers for computer calculations.

A. Autosomal dominant inheritance

Autosomal dominant inheritance has the following characteristics: (i) affected individuals are directly related in one or more successive generations, (ii) both males and females are affected in a 1:1 ratio, and (iii) the expected proportion of affected and unaffected offspring of an affected individual is 1:1 (0.50, or 50%) each (**1**). Depending on the type of autosomal disease, the first affected individual will be the result of a new mutation (**2**, **3**). In some autosomal dominant disorders, a carrier of the mutation does not manifest the disease. This is called *nonpenetrance*. The degree of manifestation can vary within a family. This is called *variable expressivity*.

B. Autosomal recessive inheritance

Autosomal recessive inheritance has the following characteristics: (i) affected siblings of both sexes in a ratio of 1:1 and (ii) unaffected parents. The genotypes of children of heterozygous patients are distributed in a ratio 1:2:1 (25% homozygous normal, 50% heterozygous, and 25% homozygous affected). Thus, heterozygous parents have a risk of 25% of affected offspring.

In the examples shown, in pedigree **1,** the unaffected parents (II-3 and II-4) must be heterozygous; the same holds true for I-1 and I-2 of pedigree **2**. In pedigree **3**, the ancestors of the affected child (IV-2) can be traced back to a common ancestor (either I-1 or I-2) of the parents (III-1 and III-2), who are first cousins (parental consanguinity, blood relationship). The double line in the pedigree indicates the parental consanguinity.

C. X-chromosomal inheritance

Since males have only one X chromosome, all daughters always inherit the father's X chromosome in addition to one of the two X chromosomes of their mother. Sons inherit the paternal Y chromosome and one of the maternal X chromosomes. Thus, the characteristic pattern of X-chromosomal inheritance is: (i) only males are affected, (ii) affected males inherit the mutant allele from the mother only, (iii) no male-to-male transmission. Therefore, a female who is heterozygous for an X-chromosomal mutation has a 50% risk of an affected son. Heterozygous females may show mild manifestations of an X-linked disease as a result of incomplete X-inactivation (see p. 184). In isolated cases with an X-chromosomal disorder, it is difficult to decide whether it results from transmission of the mutant allele from the mother or from a new mutation. Therefore, it is important to distinguish these three situations.

Females with an affected son and an affected brother or with two affected sons must be heterozygous and are said to be *obligate heterozygotes*. Those who may or may not be heterozygous are *facultative heterozygotes* (e.g., III-5 and IV-2).

Medical relevance

Human diseases following one of these inheritance patterns are referred to as *Mendelian disorders* (for a complete list see OMIM).

Further Reading

Griffith AJF, Wessler SR, Carroll SB, Doebley JF. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2015

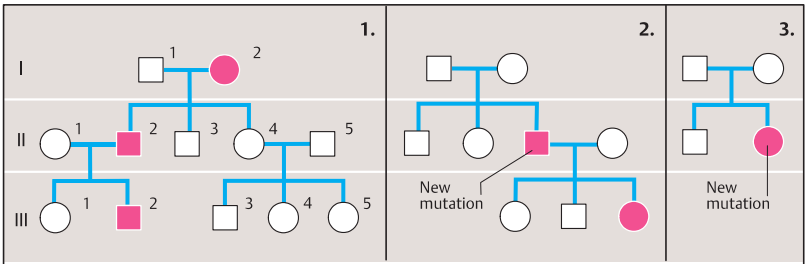
Harper PS. Practical Genetic Counselling. 7th ed. London: Edward Arnold, 2010

Jameson JL, Kopp P. Principles of human genetics In: Longo DL, et al, eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 486–509

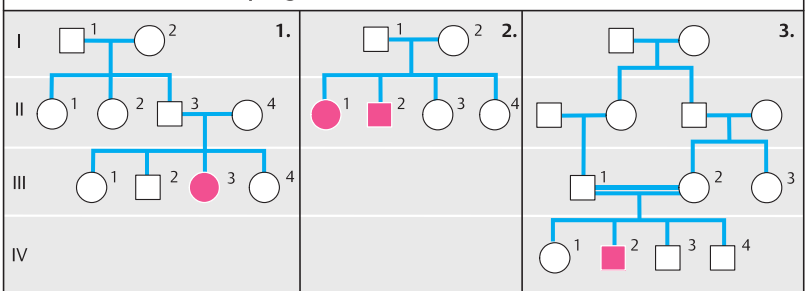
OMIM. Online Mendelian Inheritance of Man. Available at http://www.ncbi.nlm.nih.gov/omim

Rimoin DL, Pyeritz R, Korf B, eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. Philadelphia, PA: Elsevier-Churchill Livingstone, 2013 (e-book)

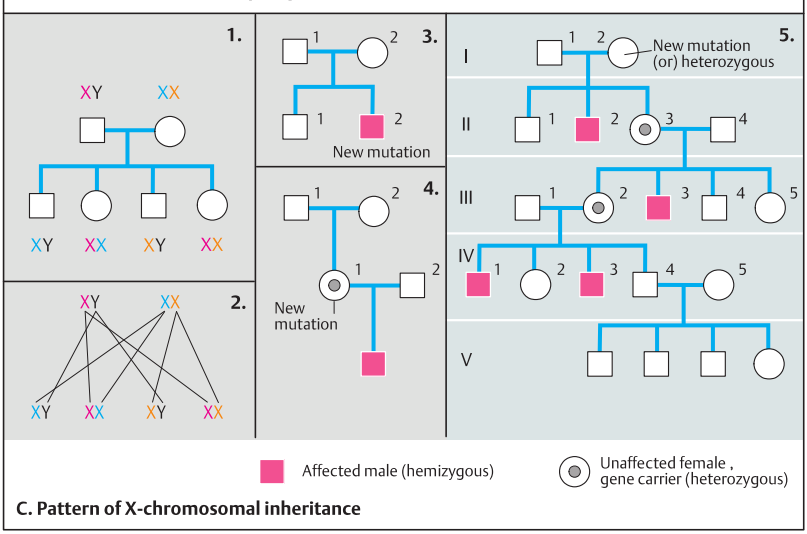
Speicher M, Antonarakis SE, Motulsky AG, eds. Vogel and Motulsky's Human Genetics. Problems and Approaches. 4th ed. Heidelberg: Springer-Verlag, 2010



A. Pattern of inheritance in pedigrees with an autosomal dominant trait



B. Pattern of inheritance in pedigrees with an autosomal recessive trait



C. Pattern of X-chromosomal inheritance

Genetic Linkage and Recombination

Genetic linkage refers to an exception from independent assortment of genetic traits. It is based on the difference whether two or more genes located near each other on the same chromosome or not. If linked, they are often transmitted together, depending how far apart they are located. Linkage was first reported in 1902 by Correns and then established by Bateson and coworkers in plant genes. Gene loci located on the same chromosome and inherited together are called a *haplotype*. Linkage relates to gene loci, not to specific alleles. *Synteny* (H. J. Renwick in 1971) refers to gene loci being located on the same chromosome without regard to linkage or the distance between them.

A. Detection of recombination

Recombination results from a cytological event (**1**) that results in genetic result (**2**).

Whether neighboring genes on the same parental chromosome remain together (linked) or become separated (recombined) depends on whether or not crossing-over (see meiosis, p. 94) between two gene loci has occurred. If the parental chromosomes harbor two allele pairs A/a and B/b, their distribution to the next generation may differ. Without crossing-over, the genotypes remain unchanged (nonrecombinant), and with crossing-over, they are rearranged (recombinant).

The genetic result (**2**) for two neighboring gene loci **A** and **B** on the same chromosome is one of two possibilities: nonrecombinant (gametes correspond to parental genotype) or recombinant (new combination). These two possibilities can be distinguished only if the parental genotypes are heterozygous (**Aa** and **Bb**), that is, are informative.

B. Use of genetic linkage

Genetic linkage has been widely used to assess the possible presence or absence of a disease-causing allele that cannot be determined directly (indirect genetic diagnosis).

If the presence of an autosomal recessive or X-chromosomal disease-causing mutation (disease locus) cannot be recognized, it can be deduced from one or several linked loci (marker locus). One uses polymorphic alleles of one or several marker loci to assess the genotype at the disease locus. Two pedigrees are shown, one without recombination (**1**) and one with recombination (**2**).

In the first example (**1**), the father and three children (red symbols in the pedigree) are affected (the children are shown as diamonds; their genders are ignored). Since all three affected children share the marker allele **A** with each other and their father, this indicates that the allele **B** at the disease locus must be the mutant allele. The unaffected children share the paternal marker locus **a**. This would indicate that the disease locus **b** is not the mutant allele. Thus, recombination has not occurred. The situation differs if recombination has occurred, as in two individuals shown (**2**). An affected individual has inherited alleles **a** and **B** from the father, instead of **A** and **B**; an unaffected individual has inherited allele **A** and allele **b**. This situation can only be observed if the father (affected parent) is heterozygous at the marker locus (**Aa**).

Medical relevance

DNA analysis of polymorphic marker loci linked to a disease locus can be used within a family to determine whether unaffected family members have inherited the mutant allele known to be present in an affected family member (indirect DNA genetic analysis). Today, usually direct recognition is possible by molecular genetic analysis.

Further Reading

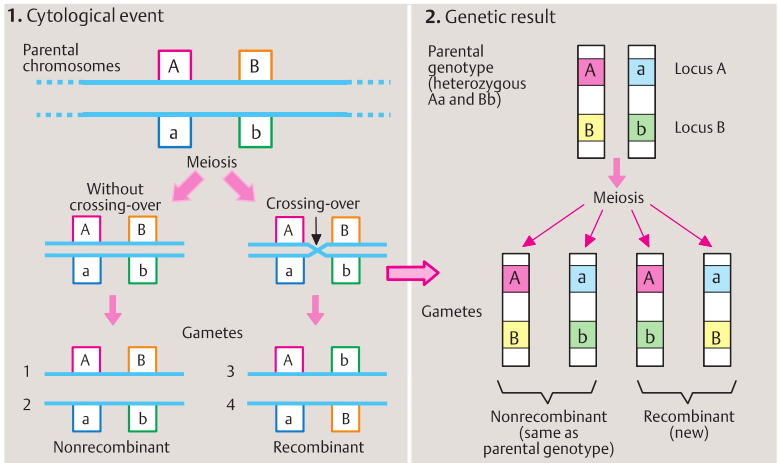
Bateson W, et al. Experimental studies in the physiology of heredity. Reports Evo-lut Comm Royal Soc 1906; 3: 1–53

Correns C. Scheinbare Ausnahmen von der Mendelschen Spaltungsregel für Bastarde. Ber deutsch bot Ges 1902; 20: 97–172

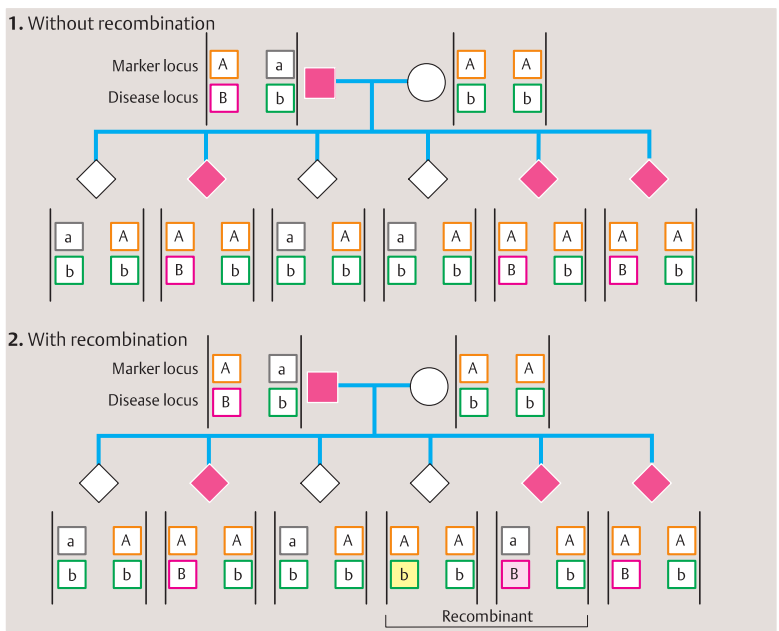
Griffith AJF, et al. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2007

Harper PS. Practical Genetic Counselling. 7th ed. London: Edward Arnold, 2010

Speicher M, et al., eds. Vogel and Motulsky's Human Genetics. Problems and Approaches. 4th ed. Heidelberg: Springer-Verlag, 2010



A. Recombination



B. Genetic linkage between two adjacent gene loci (marker locus, A, and autosomal dominant mutation, B)

Genetic Linkage and Association Analysis

Quantitative analysis of genetic linkage estimates the distance between two or more loci or adjacent DNA regions on the same chromosome. The physical distance is expressed in number of base pairs, and the genetic distance by the frequency of recombination. In contrast, genetic association expresses a statistical relation between specific alleles and a phenotype. Some alleles at different loci are inherited together more frequently or less frequently than expected by their individual frequencies. This is called *linkage disequilibrium*.

A. LOD score in linkage analysis

Quantitative genetic linkage analysis determines how often two or more gene loci are transmitted together compared with how often they are separated by recombination. This ratio is *called recombination fraction*, designated by the Greek letter theta (Θ). It expresses the genetic distance. Loci that are not linked have a Θ value of 0.5. If Θ is zero, the loci are identical. Linkage of two gene loci is assumed when the probability of linkage divided by the probability of no linkage is equal to or greater than the ratio of 1,000:1 (10^3:1). The logarithm of this ratio, the odds, is called the *LOD score*.

The four LOD score curves (**a**, **b**, **c**, **d**) are a graphic representation of linkage analysis: no linkage in **a**; recombination fraction Θ of 0.15 in **b**, that is, 15% recombination; unlikely in **c**; and inconclusive in **d**. (Figure adapted from Emery, 1986).

B. Use of haplotypes in association analysis

Genetic association analysis determines whether certain alleles are inherited more frequently than exacted by their individual frequency in the population. A deviation is referred to as *linkage disequilibrium* (LD). It can be detected by a set of associated polymorphic DNA variants (single nucleotide polymorphisms [SNPs]; see p. 72), called a *haplotype*.

Three SNP markers along a short stretch of DNA (**1**) from four unrelated individuals are shown (chromosomes 1–4). The three SNPs differ by their polymorphic alleles (A/G, T/C, C/G, respectively). This defines four haplotypes (**2**). Genotyping just three SNPs (TagSNPs) identifies the haplotypes (**3**), for example, ACG defines haplotype 2. (Figure adapted from HapMap, 2003)

C. Linkage disequilibrium

LD can be graphically represented by haplotype blocks. This compares pairwise all SNPs in the DNA region studied. The example shows this for 10 SNPs (rs7190220 at the top at the left to rs9931018 at the right). The first square at the top left shows the relationship between 39 and 40 (SNPs rs7190220 and rs178023041). The bottom square displays pair 39 and 49, and so on. The LD between each pair is shown by a color reflecting the strength of LD. A red square without a number indicates complete linkage, whereas a white square points to the absence of linkage. The numbers in the squares are percentages between pairs of loci. In the example shown, 42 (SNP rs11863156) to 49 (rs9931018) form a block of LD between 86 and 100 (bright red square without a number). If such a block occurs in higher frequency in a cohort of patients compared with controls, one may assume that this region contributes to susceptibility to the disease under study. (Figure provided by Tea Berulava Essen in 2013).

Further Reading

Emery AEH. Methodology in Medical Genetics. 2nd ed. Edinburgh: Churchill Livingstone, 1986

Evseeva I, et al. Linkage disequilibrium and age of HLA region SNPs in relation to classic HLA gene alleles within Europe. Eur J Hum Genet 2010; 18 (8): 924–932

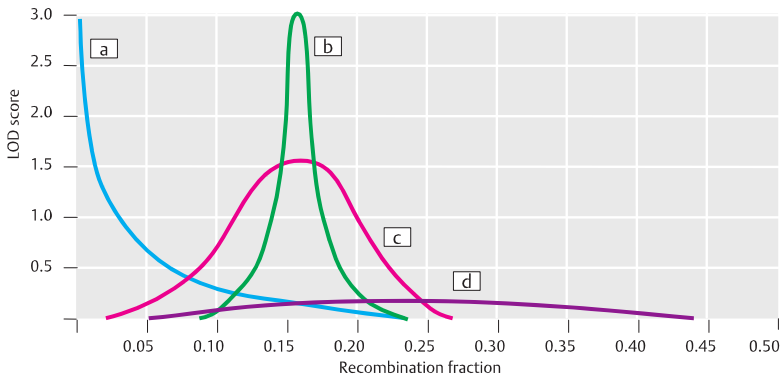
International HapMap Consortium. The International HapMap Project. Nature 2003; 426(6968): 789–796

Ott J. Analysis of Human Genetic Linkage. 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1999

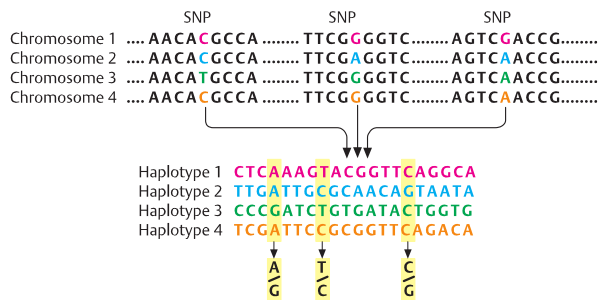
Ott J, et al. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet 2015; 16(5): 275–284

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

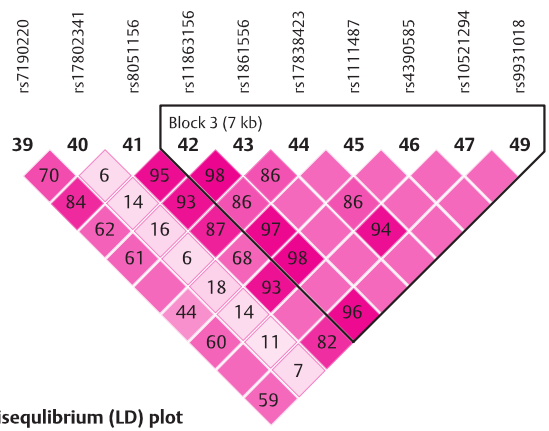
Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015



A. LOD score curves for close (a), probable (b), and no linkage (c,d)



B. Use of haplotypes in association analysis



C. Linkage disequilibrium (LD) plot

Quantitative Differences in Genetic Traits

Many phenotypes are not determined by qualitative differences rather than by measurements, for example, body height, blood pressure, and other quantitative traits. In addition, interaction with environmental factors may occur. They are not determined at a single gene locus, but by two, three, or more loci (each locus is referred to as quantitative trait locus). This is referred to as multifactorial or multigenic (polygenic) inheritance. Quantitative characteristics have a continuous distribution among different individuals. The transmission of the underlying genes cannot be recognized individually. Francis Galton (1822–1911) in 1883 introduced the concept *quantitative genetics* in contrast to Mendelian genetics, which is based on qualitative differences of phenotypes.

A. Length of the corolla in Nicotiana longiflora

The plant *Nicotiana longiflora* is an example of how the genetic transmission of a quantitative trait can be analyzed experimentally. When parent plants with average corolla lengths of 40 and 90 cm are crossed, the next generation (F_1) shows a corolla length distribution that is longer than that of the short parent and shorter than that of the tall parent. In the next generation (F_2) the distribution spreads at both ends toward long and short. If plants from the short, middle, and long varieties are crossed, the mean distribution in the next generation (F_3) corresponds to the corolla length of the parental plants. They are short (shown on the left), long (right), or average (in the middle). This can be explained by a difference in the distribution of genes contributing to the variation in the trait. (Figure adapted from Ayala & Kieger, 1984)

B. Influence of the number of gene loci on a quantitative trait

Four situations with 1, 2, 3, and 4 loci and two allele pairs **A** and **a** are shown for the parental generation **P**. The hypothetical genotypes **aa** and **AA**, **aabb** and **AABB**, **aabbcc**/**AABBCC**, and so on appear uniform in the F_1 generation. In the F_2 generation, the number of distinguishable groups depends on the number of genes involved (one in **1**, two in **2**, three in **3**, four or more in **4**). With four loci or more, the difference in size between the groups is small. It begins to resemble a continuous distribution, as shown in the bottom of the figure on the right. The smooth distribution curve corresponds to a bell-shaped normal distribution or Gaussian curve. The variance of the phenotype (V_P) is the sum of the genetic variance (V_G) and the environmental variance (V_E). The ratio V_G/V_T is called *heritability*. However, in many cases, these types of variance cannot be distinguished, especially in humans.

Medical relevance

Many relatively frequent diseases and reduced health conditions are determined by an interaction of quantitative trait loci and additional environmental factors. Examples of common quantitative genetic characteristics in humans are height, weight, eye and skin color, blood pressure, plasma concentrations of glucose and lipids, intellectual abilities, behavioral patterns, and others. Diabetes mellitus, high blood pressure, coronary heart disease, psychiatric disorders, and some congenital malformations are disease examples.

For such disorders, the risk of reoccurrence among siblings or in offspring does not follow the Mendelian patterns. Rather, this must be determined empirically in population studies.

Further Reading

Ayala FJ, Kiger JA. Modern Genetics. 2nd ed. Menlo Park, CA: Benjamin/Cummings, 1984

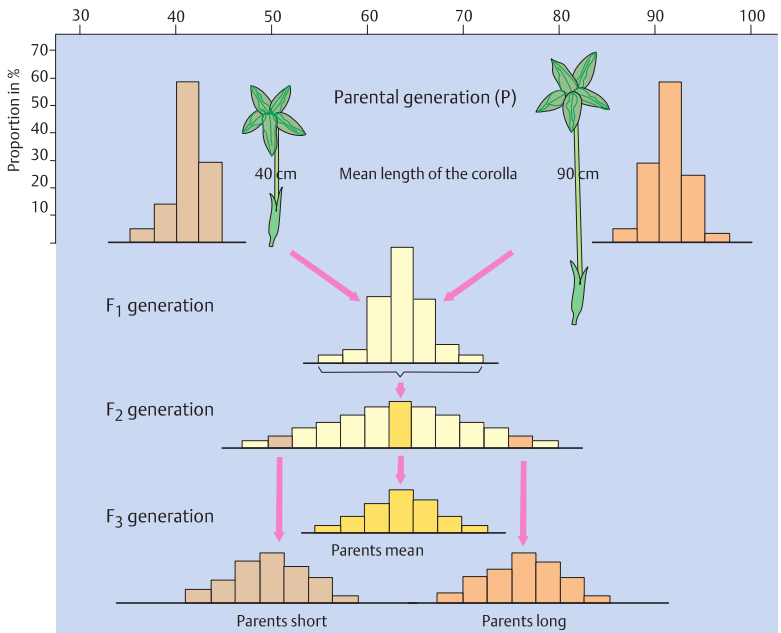
Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. 4th ed. London: Longman, 1996

Griffith AJF, Wessler SR, Carroll SB, Doebley JF. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2015

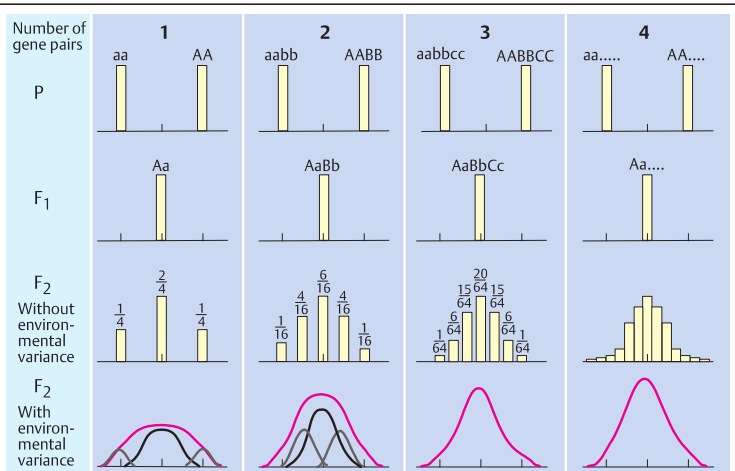
King R, Rotter J, Motulsky AG, eds. The Genetic Basis of Common Disorders. 2nd ed. Oxford: Oxford University Press, 2002

Mackay TFC. The genetic architecture of quantitative traits. Annu Rev Genet 2001; 35: 303–339

Speicher M, Antonarakis SE, Motulsky AG, eds. Vogel and Motulsky's Human Genetics. Problems and Approaches. 4th ed. Heidelberg: Springer-Verlag, 2010



A. Length of the corolla of *Nicotiana longiflora* after a cross between different types of parental plants



B. Distribution of frequency in the F_2 generation with a different number of gene loci

Distribution of Alleles in a Population

The distribution of genotypes possible among the offspring of different parental genotypes considered on p. 112 differs according to their relative frequency in the population. This is part of the field of population genetics, the scientific study of the genetic composition of populations. A principal goal is to estimate the frequency of alleles at different gene loci in natural populations (*allele frequency*, also called *gene frequency*). From this, conclusions can be drawn about possible selective influences that might explain differences observed. A population can be characterized on the basis of the frequency of certain alleles at various gene loci. Today this knowledge is obtained by sequencing the whole genome from many individuals of a population.

A. Frequency of genotypes in the children of parents with various genotypes

With regard to an allele pair A (dominant) and a (recessive), six types of parental genotype matings are possible (**1–6**). Each of these has an expected distribution of genotypes in the offspring according to the Mendelian laws, as indicated in the figure. This pattern will only be observed if all genotypes can participate in mating and are not prohibited by a severe disease. The frequency with which each mating type occurs depends on the frequencies of the alleles in the population.

B. Allele frequency

The frequency of allele designates the proportion of a given allele at a given locus in a population. If an allele accounts for 20% of all alleles present (at a given locus) in the population, its frequency is 0.20. The allele frequency determines the frequencies of the individual genotypes in a population.

For example, for a gene locus with two possible alleles **A** and **a**, three genotypes are possible: **AA**, **Aa**, or **aa**. The frequency of the two alleles together (p the frequency of **A**, and q the frequency of a) is 1.0 (100%). If two alleles, **A** and **a**, are equally frequent (each 0.5), they have the frequencies of p = 0.5 for allele **A** and q = 0.5 for allele **a** (**1**). Thus, the equation $p + q = 1$ defines the population at this locus. The frequency distribution of the two alleles in a population follows a simple binomial relationship: $(p + q)^2 = 1$. Accordingly, the distribution of genotypes in the population corresponds to $p^2 + 2pq + q^2 = 1.0$. The expression p^2 corresponds to the frequency of the genotype **AA**, the expression $2pq$ corresponds to the frequency of the heterozygotes **Aa**, and q^2 corresponds to the frequency of the homozygotes **aa**. This relationship is called the Hardy-Weinberg principle.

When the frequency of an allele is known, the frequency of the genotype in the population can be determined as follows: given a frequency p of allele **A** of 0.6 (60%), the frequency q of allele **a** is 0.4 (40%, derived from $q = 1 - p$ or 1 – 0.6). Thus, the frequency of the genotype **AA** is 0.36, that of **Aa** is 2 × 0.24 = 0.48, and that of **aa** is 0.16 (2). If only the homozygotes **aa** can be recognized (e.g., an autosomal recessive inherited disease) but not the heterozygotes Aa, then q^2 (0.16 in this example) corresponds to the frequency of the disorder. From $p = 1 - q$, the frequency of heterozygotes ($2pq$) and of normal homozygotes (p^2) can also be determined.

Medical relevance

In genetic counseling, the genotype of an unaffected sibling of an individual with an autosomal recessive disorder can be estimated from the genotype distribution 1:2:1. Since the unaffected sibling cannot be homozygous for the mutant allele, the probability of heterozygosity would be ⅔ (66%), derived from the remaining ratio of 2:1.

Further Reading

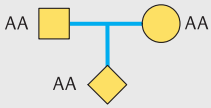
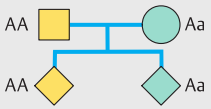
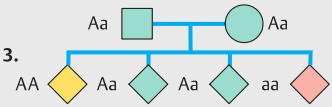
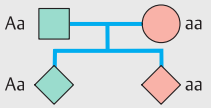
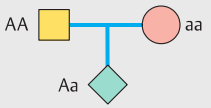
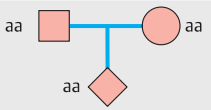
Cavalli-Sforza LL, Bodmer WF. The Genetics of Human Populations. San Francisco, CA: W. H. Freeman, 1971

Hamilton MB. Population Genetics. Chichester: John Wiley & Sons, 2009

Kimura M, Ohta T. Theoretical Aspects of Population Genetics. Princeton, NJ: Princeton University Press, 1971

Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999; 22(2): 139–144

Speicher M, et al, eds. Vogel and Motulsky's Human Genetics. Problems and Approaches. 4th ed. Heidelberg: Springer-Verlag, 2010

		Genotype of parents	Genotype of offspring
1.		AA and AA	1.0 AA
2.		AA and Aa	0.50 AA 0.50 Aa
3.		Aa and Aa	0.25 AA 0.50 Aa 0.25 aa
4.		Aa and aa	0.50 Aa 0.50 aa
5.		AA and aa	1.0 Aa
6.		aa and aa	1.0 aa

A. Expected frequency of genotypes in children of parents with different genotypes

1.

Parents	0.5 A	0.5 a
0.5 A	AA 0.25	Aa 0.25
0.5 a	Aa 0.25	aa 0.25

Offspring

$p = 0.50$ (Frequency of A)
$q = 0.50$ (Frequency of a)

2.

	A = 0.60	a = 0.40	
A 0.6	0.36 AA	0.24 Aa	p
a 0.4	0.24 Aa	0.16 aa	q
	p	q	

$p^2 + 2pq + q^2 = 1$
$0.36 + 0.48 + 0.16 = 1.0$
(AA) (Aa) (aa)

B. Allele frequency

Hardy–Weinberg Equilibrium Principle

The Mendelian segregation of alleles under random mating results in the frequency of genotypes being constant in every new generation. This is the Hardy–Weinberg equilibrium principle, first formulated independently by the English mathematician G. F. Hardy and the German physician W. Weinberg in 1908. However, if mating is not random, the frequencies will increase or decrease.

A. Constant allele frequency

An autosomal recessive allele (allele **a**) that leads to a severe disorder in the homozygous state remains undetectable in the heterozygous state in a population. Only the homozygotes (**aa**) can be recognized because of their disease. The frequency of affected individuals (homozygotes **aa**) depends on the frequency of allele **a** in the population (corresponding to q). The frequency of the three genotypes is determined by the binomial relationship $(p + q)^2 = \mathbf{1}$, where p represents the frequency of allele **A**, and q represents the frequency of allele **a** (see previous page). The homozygous alleles (**aa**) eliminated in one generation by illness are replaced by new mutations. This results in an equilibrium between elimination due to illness and frequency of the mutation.

B. Some factors influencing the allele frequency

The Hardy–Weinberg equilibrium principle is valid only in certain conditions. First, it applies only if there is no selection for one genotype (random mating). Selection for heterozygotes will increase the frequency of the allele with a selective advantage (see next page). Second, nonrandom matings (assortative mating) will change the allele frequency (proportion of p and q). Third, a change in the rate of mutations will increase the frequency of the allele resulting from mutations. Fourth, in a small population, random fluctuation may change the frequency. This is called *genetic drift*.

Other causes of a change in allele frequency may occur. If a population experiences a drastic reduction in size (genetic bottle neck), followed by a subsequent increase in the number of individuals, an allele that was previously rare in this population may, by chance, subsequently become relatively common as the population expands again. This is called a *founder effect*. Photograph from Coney Idland 1938, Weegee, also known as Arthur Fellig).

Medical relevance

The frequency of heterozygotes for an autosomal recessive disorder can be derived from the Hardy–Weinberg equation ($p + q = 1$). If, for example, the disease frequency is 1:10,000 (0.01%), then $q^2 = 1/10{,}000$ would yield $q = 1/100$. Since $2pq$ corresponds to the heterozygote frequency, this would be 1:50 since $2pq$ is $2 \times 1 \times 1/100$. A founder effect in some rare autosomal recessive disorders has frequently been observed in closed human populations.

Further Reading

Cavalli-Sforza LL, Bodmer WF. The Genetics of Human Populations. San Francisco, CA: W. H. Freeman, 1971

Cavalli-Sforza LL, et al. The History and Geography of Human Genes. Princeton, NJ: Princeton University Press, 1994

Hardy GH. Mendelian proportions in a mixed population. Science 1908; 28(706): 49–50

Jorde LB. Linkage disequilibrium and the search for complex disease genes. Genome Res 2000; 10(10): 1435–1444

Kimura M, Ohta T. Theoretical Aspects of Population Genetics. Princeton, NJ: Princeton University Press, 1971

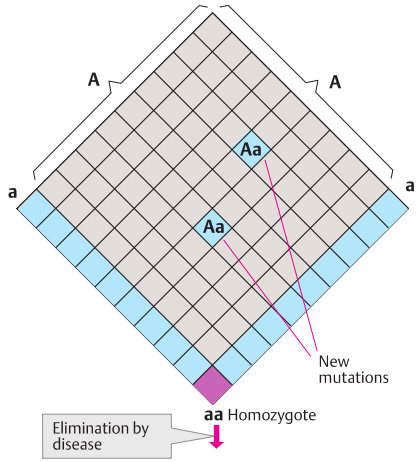
Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999; 22(2): 139–144

Speicher M, et al, eds. Vogel and Motulsky's Human Genetics. Problems and Approaches. 4th ed. Heidelberg: Springer-Verlag, 2010

Weinberg W. Über den Nachweis der Vererbung des Menschen. Jahreshefte Verein vaterländ Naturk Württemberg 1908; 64: 368–382

Wigginton JE, et al. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005; 76 (5): 887–893

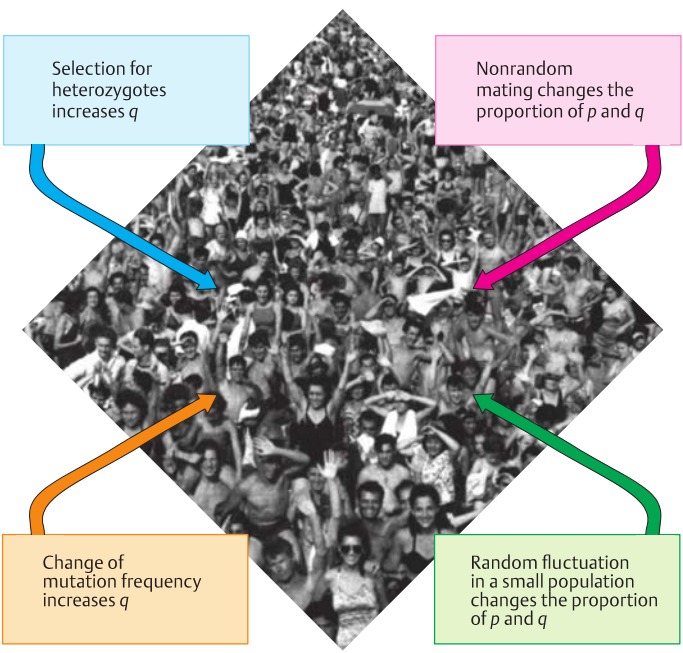
Zöllner S, von Haeseler A. Population History and Linkage Equilibrium. Nature Encyclopedia of the Human Genome. Vol. 4. London: Nature Publishing Group, 2003: 628–637



Alleles **a** are eliminated from the population owing to severe illness in homozygotes **aa**, but are replaced by new mutation. Equilibrium results

Genotypes		Frequency
Homozygote	**AA**	p^2
Heterozygote	**Aa**	$2pq$
Homozygote	**aa**	q^2

A. Constant allele frequency (Hardy–Weinberg equilibrium)



B. Some factors influencing the allele frequency

Geographical Differences in Allelic Distribution

Human populations from different geographical regions differ in the frequency of many alleles. This is a result of migration of human ancestors to different parts of the world during human evolution. This genetic diversity is defined by differences in the frequency of DNA nucleotide sequences and the frequency of certain alleles at various gene loci. Certain mutant alleles may differ widely in frequency. Therefore, autosomal recessive diseases that are common in one population may be rare in others. This may either be the result of random fluctuations of allele frequency of some alleles or a selective advantage of an allele.

A. Different frequencies

Finland represents an example of a small population in which several rare recessive diseases occur at a much higher frequency than elsewhere. The most likely explanation is a combination of founder effect and genetic isolation. Three diseases that are clustered in different regions of Finland are (**1**) congenital flat cornea (cornea plana 2, OMIM 217300) in the western part of the country, (**2**) the Finnish type of congenital nephrosis, a severe renal disorder (OMIM 256300) in the southwestern part, and (**3**) diastrophic skeletal dysplasia (OMIM 222600) in the southeastern region. The recessive mutant alleles in these diseases are identical by descent. They must have arisen independently in the different regions because the grandparents live predominantly there. There is no known basis for a selective advantage of these mutant alleles in the heterozygotes to explain their frequencies. The different distributions merely reflect the places and relative points in time of the mutations. Similar examples are found in many other regions of the world and in other populations. (Figures from Norio, 1981).

B. Malaria and hemoglobin disorders

The distribution of the parasitic disease, malaria, overlaps closely with the distribution of different types of hemoglobin disorder (see p. 350). Malaria is common in tropical and subtropical regions (**1**). It used to be common around the Mediterranean Sea, but here the frequency of malaria has been reduced as a result of successful prevention. In the same regions with endemic malaria, several types of hemoglobin disorders are prevalent (**2, 3**). Typical examples are sickle cell anemia (OMIM 141900) and different types of thalassemia (OMIM 187550). In 1954, A. C. Allison proposed that individuals who are heterozygous for the sickle cell mutation are less susceptible to malaria infection. This is the first and the best example of a heterozygous selective advantage known in man.

The sickle cell mutant has arisen independently at least four times in different regions and has become established as a result of selective advantage for heterozygotes.

Another red blood cell disease confers an advantage against malaria infection on heterozygotes: glucose-6-phosphate dehydrogenase deficiency (OMIM 305900), an X-chromosomal anemia disorder leading to severe anemia in hemizygous males, whereas heterozygous females are normal and relatively protected against malaria.

The selective advantage of heterozygotes for these genetically determined diseases is based on less favorable conditions for the malaria parasite than in the blood of normal homozygotes. The protection of heterozygotes occurs at the expense of affected homozygotes who suffer from one of the severe hemoglobin disorders. The benefits are at the population level.

Further Reading

Allison AC. The distribution of sickle-cell trait in East Africa and elsewhere, and its apparent relationship to the incidence of subtertian malaria. Trans R Trop Med Hyg 1954; 48(4):312–318

Cavalli-Sforza LL, et al. The History and Geography of Human Genes. Princeton, NJ: Princeton University Press, 1994

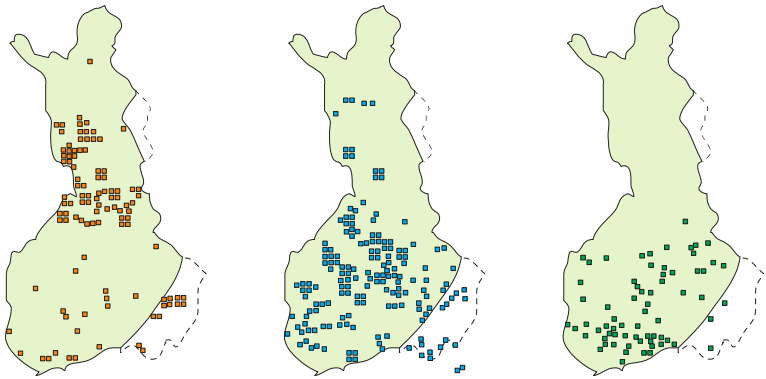
Greenwood BM, et al. Malaria. Lancet 2005; 365 (9469): 1487–1498

Marshall E. Malaria. A renewed assault on an old and deadly foe. Science 2000; 290(5491): 428–430

Norio R. Diseases of Finland and Scandinavia. In: Rothschild HR, ed. Biocultural Aspects of Disease. New York, NY: Academic Press, 1981

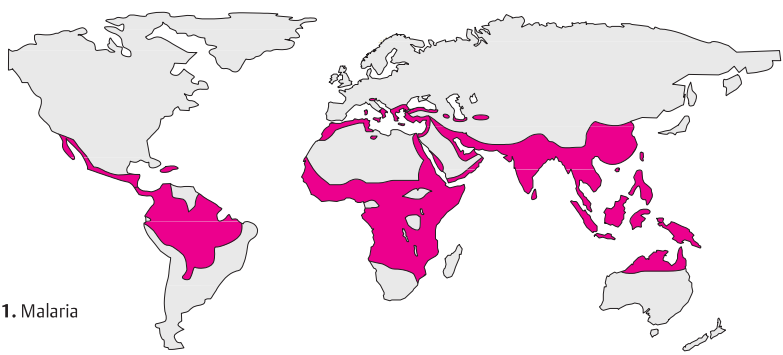
Norio R. The Finnish Disease Heritage III: the individual diseases. Hum Genet 2003; 112(5-6): 470–526

Weatherall DJ, Clegg JB. Thalassemia—a global public health problem. Nat Med 1996; 2(8): 847–849

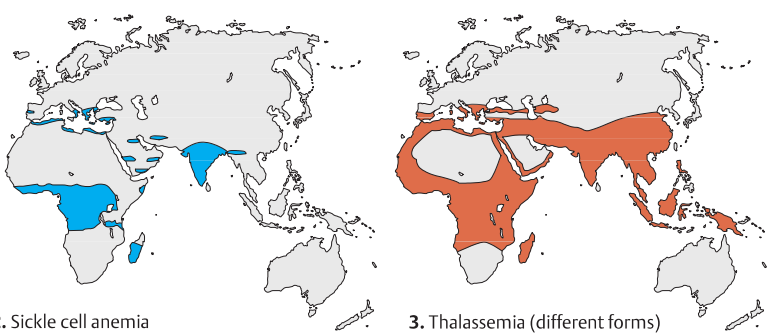


1. Congenital corneal dystrophy **2.** Congenital nephrosis **3.** Diastrophic skeletal dysplasia

A. Different frequencies of genetic diseases, e.g., in Finland



1. Malaria



2. Sickle cell anemia **3.** Thalassemia (different forms)

B. Distribution of malaria and hemoglobin diseases

Inbreeding

Inbreeding in humans refers to matings between individuals who are closely related to each other. If the parents have at least one common ancestor in the past two to four generations, there is an increase in the chance that an allele present in the heterozygous state in both parents will become homozygous in a descendant. This is referred to as parental consanguinity (being of the "same blood"). As both alleles will be identical, this is called *identity by descent* (IBD). The degree of consanguinity can be expressed by two measures: coefficient of relationship (r) and the inbreeding coefficient (F). The coefficient of relationship (r) expresses the degree of consanguinity between two individuals. The inbreeding coefficient expresses the probability that two alleles will be homozygous in a population.

A. Simple types of consanguinity

A mating between brother and sister or father and daughter is called *incest* (**1**). For two alleles descending from both parents, here designated as A and B, the probability of transmission to each of the two offspring, C and D, respectively, is 0.5 (1/2). The siblings (C and D) share half of their genes, corresponding to a coefficient of relationship of 1/2. The chance of homozygosity by descent at a given locus in their offspring is 1/4 (1/2 each from C and D to E).

First cousins share 1/8 of their genes (**2**), and second cousins share 1/32 of their genes. The chance of homozygosity by descent in the offspring of first cousins is 1/16. An uncle–niece union (**3**) has a coefficient of relationship of r = 1/4 because they share 1/4 of their genes. The chance of homozygosity by descent in their offspring is 1/8. The possibility that an unrelated individual E transmits a mutant allele at this locus can usually be disregarded.

B. Identity by descent

IBD refers to regions in the genome that are identical because they are derived from a common ancestor. For example, for two alleles A and B in ancestor I, and C and D in ancestor II, the probability of transmission to the next generation is 0.5 (1/2) for each. The same applies to the next generation with individuals III and IV. Finally, the probability of transmission from III to V, and IV to V, is again 0.5 or 1/2 each. The number of steps (the probability of transmission) from I to III, and from I to IV, is $(1/2)^2$ each. However, for individual V to be homozygous, it is required that both alleles are transmitted from I to III and to IV. This corresponds to $(1/2)^4$. For individual V to be homozygous by IBD, the probabilities $(1/2)^4$ have to be added and multiplied by the probability of (1/2), which yields 1/16.

Medical relevance

Consanguinity favors the occurrence of homozygosity for a disease-causing allele present in heterozygous parents. First cousins with a coefficient of relationship r of 1/8 (0.125) have a statistical risk for homozygous offspring of 0.312 (3.12%). Although at first, this risk seems high, it is not when compared with the risk in the general population. The overall risk for a newborn to have a disorder of any kind is estimated to be 1 to 2%. Although consanguineous marriage is widespread in some populations, where it may account for 25 to 40% of marriages, the usual rate is approximately 1 to 2%.

Further Reading

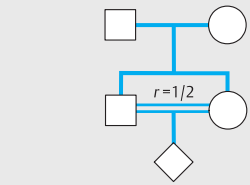
Bittles AH, Neel JV. The costs of human inbreeding and their implications for variations at the DNA level. Nat Genet 1994; 8(2): 117–121

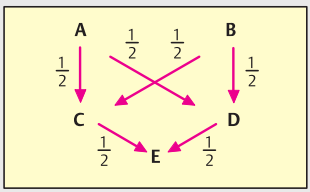
Griffith AJF, et al. An Introduction to Genetic Analysis. 10th ed. New York, NY: W. H. Freeman, 2015

Harper PS. Practical Genetic Counselling. 7th ed. London: Edward Arnold, 2010

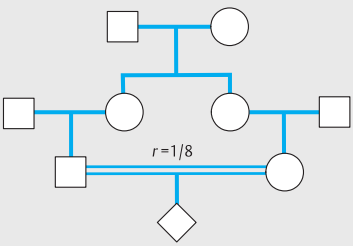
Jaber L, et al. The impact of consanguinity worldwide. Community Genet 1998; 1(1): 12–17

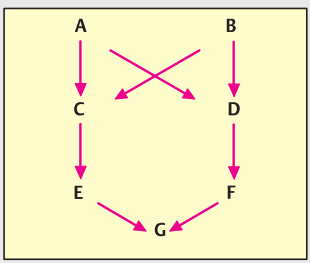
Turnpenny P, Ellard S. Emery's Elements of Medical Genetics. 14th ed. Edinburgh: Elsevier-Churchill Livingstone, 2011



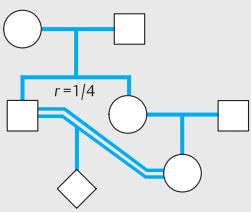


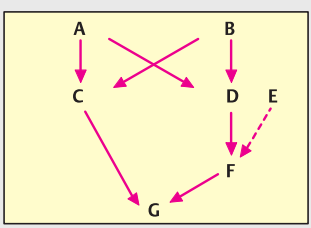
1. Brother/sister mating





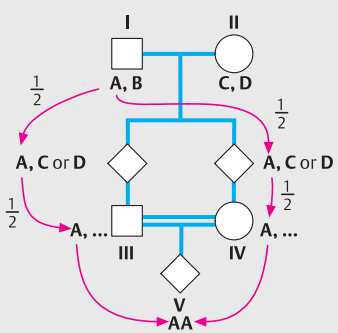
2. First cousins

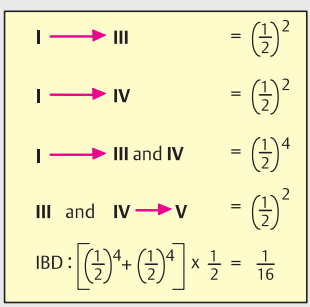




3. Uncle/niece

A. Simple types of consanguinity





Probability of descent from a common ancestor

B. Identity by descent (IBD)

Twins and Twinning

In humans, twinning is detected during pregnancy by ultrasonography in approximately 1 of 40 pregnancies, but in live births in 1 of 80. This difference results from the early intrauterine death of one twin and its subsequent resorption (vanishing twin). The frequency of twinning varies widely, approximately 6 in 1,000 births in Asia, 10 to 20 in 1,000 in Europe, and 40 in 1,000 in Africa. Twins can arise from a single fertilized egg (genetically identical monozygotic twins, MZ) or from two different eggs (dizygotic twins, DZ), a distinction first proposed by C. Dareste in 1874. The rate of monozygotic twins at birth is relatively constant.

A. Types of human monozygotic twins

Monozygotic twins arise during very early stages of embryonic development by splitting of the inner cell mass of the early embryo. Three stages in the timing can be distinguished: (**1**) splitting after trophoblast formation, resulting in twins with individual amnions but one common chorion, (**2**) splitting after amnion formation, resulting in twins in one single chorion and amnion, and (**3**) splitting before formation of the trophoblast, resulting in twins each with its own chorion and amnion.

DZ by definition should have each one chorion and one amnion cavity. However, depending on the time of splitting, approximately 66% of MZ twins have one chorion and two amnions (splitting after formation of the chorion at day 5, but before formation of the amnion at day 9). Approximately 33% of MZ twins have two complete separate chorions and a common amnion.

B. Pathological conditions in twins

Different structural defects can arise in twins. This is three times more frequent in MZ than in DZ. The defects can be assigned to three groups: (1) incomplete or late splitting (Thoracopagus, twins joined to various extents), (2) shared blood supply by a shunt, (3) absent heart in one twin (acardius). Fetal space constraints with crowing in the uterus may lead to different deformities.

C. Concordance rates in twins

MZ and DZ differ with respect to their frequency of traits occurring in both (concordance rate). When twins show the same trait, they are said to be *concordant* for that trait; when they differ, they are *discordant*. Systematic studies of MZ and DZ twins have compared their rate of concordance in an attempt to determine the relative contribution of genetic factors in the etiology of complex traits.

D. Pharmacogenetic pattern in twins

Dizygotic and monozygotic twins also differ biochemically. As a result of genetic differences, many chemical substances used in therapy are metabolized or excreted at different rates, owing to different activities of corresponding enzymes. Phenylbutazone is excreted at the same rate in identical twins, whereas the rates of excretion differ between dizygotic twins or among siblings (Vesell, 1978).

Further Reading

Boomsma D, et al. Classical twin studies and beyond. Nat Rev Genet 2002; 3(11): 872–882

Bouchard TJ Jr, et al. Sources of human psychological differences: the Minnesota Study of Twins Reared Apart. Science 1990; 250(4978): 223–228

Gilbert SF. Developmental Biology. 9th ed. Sunderland, MA: Sinauer, 2010

Goerke K. Taschenatlas der Geburtshilfe. Stuttgart: Thieme, 2002

Hall JG. Twinning. Lancet 2003; 362(9385): 735–743

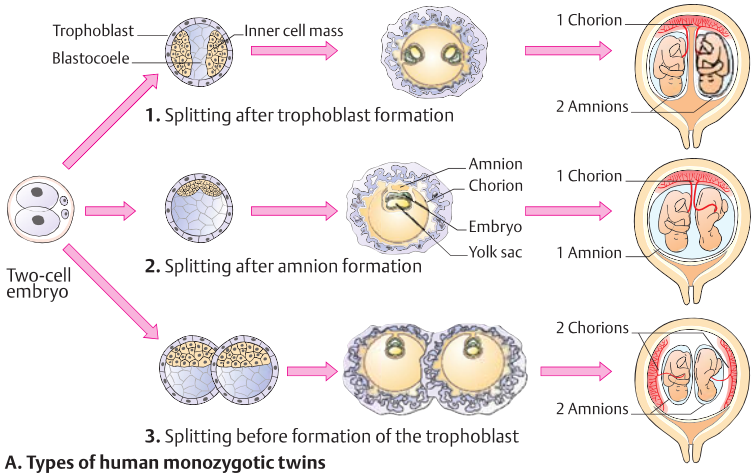
Hall JG. Twins. chapter 35. In: Stevenson RE, et al., eds. Human Malformations and Related Anomalies. 3 rd. ed. Oxford University Press, Oxford, 2015

MacGregor AJ, et al. Twins. Novel uses to study complex traits and genetic diseases. Trends Genet 2000; 16(3): 131–134

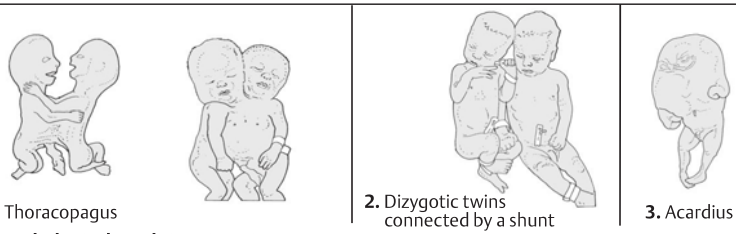
Painter JN, et al. Twins and twining. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013 (e-book only)

Stone JL, et al. The craniopagus malformation: classification and implications for surgical separation. Brain 2006; 129(Pt 5): 1084–1095

Vesell ES. Twin studies in pharmacogenetics. Hum Genet Suppl 1978; 1(1, Suppl): 19–30



A. Types of human monozygotic twins

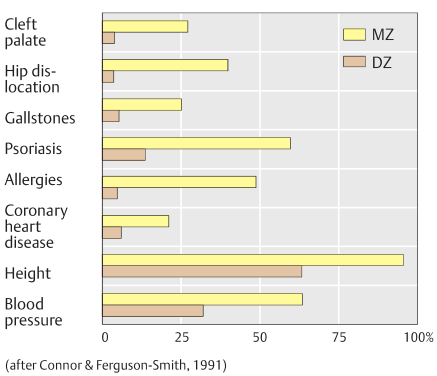


1. Thoracopagus

2. Dizygotic twins connected by a shunt

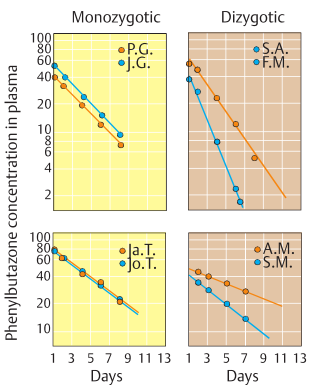
3. Acardius

B. Pathological conditions in twins



(after Connor & Ferguson-Smith, 1991)

C. Concordance of some traits in monozygotic (MZ) and dizygotic (DZ) twins



D. Pharmacogenetic pattern in twins

Chromosomes and Genes

In the early 20th century, it was realized that genetic factors, called *genes* in 1909, and threadlike darkly stained structures in the nucleus of eukaryotic cells, called *chromosomes* in 1888, came in pairs. However, the sciences of cytology and genetics, a term introduced in 1906, remained separated. It was in 1912, when Thomas H. Morgan and his coworkers at Columbia University, New York, recognized that genes are located on chromosomes in a defined order, a genetic map, as first shown by Sturtevant in 1913.

A. Polytene chromosomes

The four *Drosophila* chromosomes illustrated (chromosomes 1–3, X, and Y) show approximately 5,000 bands in a chromosome-specific pattern of lightly and darkly stained bands. A micrographic detail of a chromosome from a *Drosophila* salivary gland shows details of the characteristic banding pattern (shown at the bottom of **A**).

Polytene chromosomes (many threads) occur in certain cells and tissues of some insects and amphibians contain such chromosomal structures. They form as a result of repeated DNA synthesis without cell division. These chromosomes, first observed in cells of insect salivary glands (*Drosophila melanogaster* and *Chironomus*) by E. G. Balbiani in 1881, also show regions of temporary localized enlargement called *Balbiani rings*. (Figures adapted from Alberts et al, 2015 and Painter 1934).

B. Functional stages in polytene chromosomes

Expansions (chromosome puffs) appear and recede at defined positions and developmental stages in polytene chromosomes (**1**). The expanded segments represent regions containing active genes that are being transcribed. The locations and durations of the puffs reflect different stages of larval development. The incorporation of radioactively labeled RNA (**2**) demonstrates that RNA synthesis occurs in these regions as a sign of gene activity (transcription).

C. Lampbrush chromosomes in oocytes

In the oocytes of some animals, fine loops protrude from the chromosomes during the diplotene phase of meiosis (see p. 94). Because of their appearance, they were called *lampbrush chromosomes* by W. Flemming in 1882. These are greatly extended chromosome bivalents during the diplotene phase of meiosis (**1**). A meiotic bivalent consisting of two pairs of sister chromatids seen under the light microscope are held together at points of chiasma formation (**2**). (Microphotograph by Gall, 1956.)

D. Visible transcription of ribosomal RNA gene clusters

Tandem repeats of ribosomal RNA (rRNA) genes transcribed in the nucleolus of an amphibian, *Triturus viridescens*, are shown here. Along each gene, many rRNA molecules are synthesized by RNA polymerase I. The growing RNA molecules extend from a backbone of DNA. (Microphotograph by Miller, 1973).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Ashburner M, et al. Temporal control of puffing activity in polytene chromosomes. Cold Spring Harb Symp Quant Biol 1974; 38: 655–662

Callan HG. The Croonian Lecture, 1981. Lampbrush chromosomes. Proc R Soc Lond B Biol Sci 1982; 214 (1197): 417–448

Gall JG. On the submicroscopic structure of chromosomes. Brookhaven Symp Biol 1956; 8(8): 17–32

Griffith AJF, et al. An Introduction to Genetic Analysis. 9th ed. New York, NY: W. H. Freeman, 2007

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

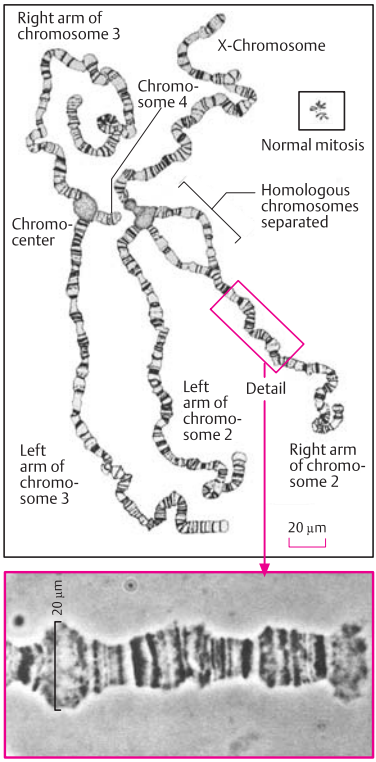
Miller OL Jr. The visualization of genes in action. Sci Am 1973; 228(3): 34–42

Morgan TH, et al. The Mechanism of Mendelian Heredity. New York, NY: H. Holt & Co, 1915

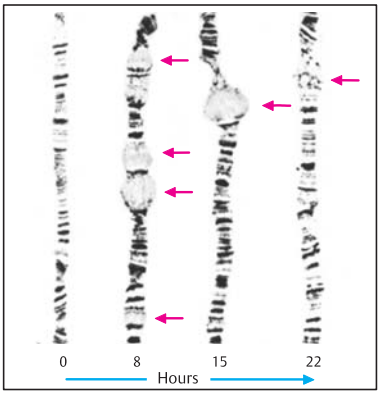
Morgan TH. The Theory of the Gene. New Haven, CT: Yale University Press, 1926

Painter TS. Salivary chromosomes and the attack on the gene. J Hered 1934; 25(12): 465–476

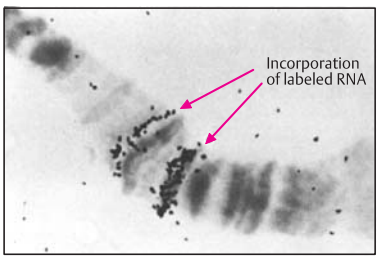
Sturtevant AH. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J Exp Zool 1913; 14(1): 43–59



A. Polytene chromosomes in salivary glands of *Drosophila* larvae

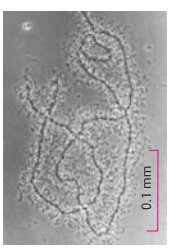


1. Formation of puffs (arrows)

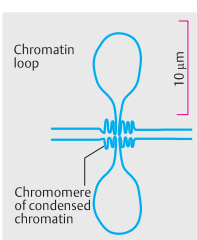


2. Evidence of gene activity

B. Functional stages in polytene chromosomes

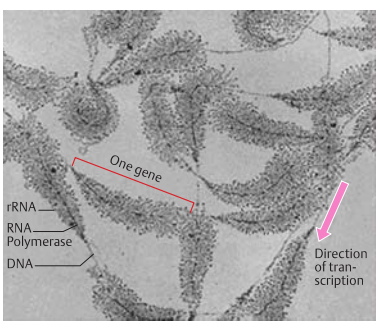


1. Lampbrush chromosome



2. Schematic section of a chromatin loop

C. Chromosome structure in amphibian oocytes ("lampbrush chromosomes")



D. Visible transcription of ribosomal RNA gene clusters

Chromosome Organization

The structural organization of chromosomes differs between interphase and mitosis. Chromosomes are visible as separate structures during mitosis only. During interphase, they appear in the cell nucleus as a tangled mass called *chromatin*. The local density of chromatin shows regions of light and dark stain through a light microscope. These are referred to as *heterochromatin* and *euchromatin* (E. Heitz, 1928). Euchromatin is related to active genes, and heterochromatin is related to inactive genes.

A. Chromosome fibers

The skeleton of a metaphase chromosome is shown in the center, surrounded by a scaffold of darkly stained dense fibers (**1**). The chromosome has been depleted of histone proteins (see p. 138). The fibers represent a halo of DNA (darkly stained threads). A higher magnification (**2**) indicates that the DNA consists of a single continuous thread. (Photographs Paulson & Laemmli, 1977).

B. Levels of chromosome organization

From the chromosome to its DNA strand, different levels of organization can be distinguished in a zoom-like manner. The total length of haploid DNA in a dividing human cell is approximately 1 m. During mitosis, this has to fit into 23 chromosomes of approximately 3 to 7 µm each. When a portion of a chromosome arm corresponding to 10% of that chromosome is magnified tenfold, that chromosomal area might contain approximately 40 genes, depending on the part of chromosome segment (eight are shown here, **2**). Magnifying a tenth of that area another tenfold (**3**) would yield a region containing three or four genes on average (3 shown). A further tenfold magnification brings this down to the level of a single gene (**4**), here shown with 7 exons (E1–E7). The last level (**5**) would be that of the nucleotide sequence.

C. Heterochromatin and euchromatin

In 1928, Emil Heitz observed that certain parts of the chromosomes of a moss (*Pellia epiphylla*) remain thickened and deeply stained during interphase. He named these structures *heterochromatin*. Those parts that were less densely stained and became invisible during late telophase and subsequent interphase he called *euchromatin*. Subsequent studies showed that heterochromatin consists of regions with few or no active genes, whereas euchromatin corresponds to regions with active genes. (Figure from Heitz, 1928).

D. Constitutive heterochromatin at the centromeres (C bands)

The centromeres of eukaryotic chromosomes contain repetitive DNA, called *α-satellites*. They are specific for each chromosome. These sequences are visible in the centromeric region (constitutive heterochromatin). This can be specifically stained (C bands). The heterochromatin of the centromeres of chromosomes 1, 9, and 16, and of the long arm of the Y chromosome, differs in length in different human individuals (chromosomal polymorphism). (Photograph Verma & Babu, 1989).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Brown SW. Heterochromatin. Science 1966; 151 (3709): 417–425

Grewal SI, et al. Heterochromatin revisited. Nat Rev Genet 2007; 8(1): 35–46

Heitz E. Das Heterochromatin der Moose. I. Jahrb Wiss Bot 1928; 69: 762–818

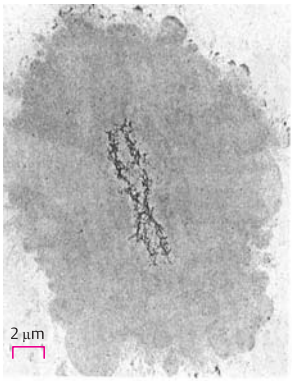
Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

Passarge E. Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. Am J Hum Genet 1979; 31(2): 106–115

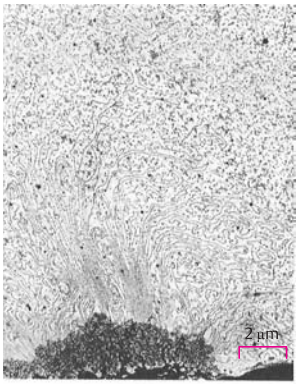
Paulson JR, Laemmli UK. The structure of histone-depleted metaphase chromosomes. Cell 1977; 12(3): 817–828

Sumner A. Chromosomes: Organization and Function. Malden, MA: Blackwell, 2003

Verma AS, Babu A. Human Chromosomes. New York, NY: Pergamon Press, 1989

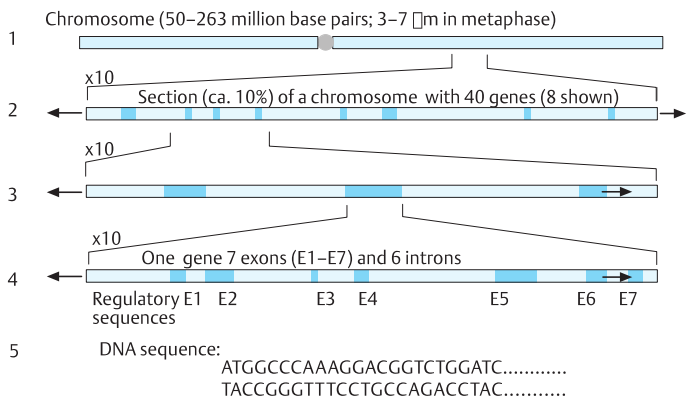


1.

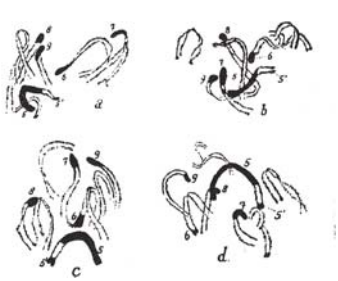


2.

A. Chromosome fibers



B. Levels of chromosome organization



C. Heterochromatin and euchromatin

D. Constitutive heterochromatin (C bands) at the centromeres

Functional Elements of Chromosomes

The eukaryotic chromosome has been viewed as a segragation device by Benjamin Lewin in 2008. Three types of structures are required to replicate and be distributed at mitosis: (i) sequences at the centromere (CEN); (ii) autonomously replicating sequences (ARSs, origins of replication); and (iii) sequences at the telomer (TEL). The individual contributions of these three types of functional chromosomal elements have been demonstrated in mutant yeast cells (*Saccharomyces cerevisiae*, baker's yeast). Using this information, yeast artificial chromosomes can be constructed.

A. Basic features of a eukaryotic chromosome

The *centromere* and the two *telomeres*, one at each end, are distinct hallmarks of a chromosome. The centromere attaches a chromosome to the spindle at mitosis. Centromeric (CEN) DNA contains repetitive DNA called *α-satellite DNA*, the most abundant type being long tandem repeats of a 170-base-pair (bp) monomeric sequence. The total length of CEN DNA ranges from approximately 300 to 5,000 kb. Cloned α-satellite fragments hybridize specifically to individual chromosomes. This can be used to identify individual chromosomes. The subtelomeric sequences, located proximal to the telomeric (TEN) sequences, contain sequence homologies shared among subsets of other chromosomes (see p. 142). Each chromosome is characterized by a specific pattern of dark and light bands (banding pattern).

B. Autonomously replicating sequences

ARSs are required for initiation of eukaryotic DNA replication. This can be shown in mutant yeast cells. Mutant cells lacking the gene for synthesis of leucine (Leu) cannot grow in culture medium that lacks leucine. Even if Leu is added, they cannot replicate because they lack ARS (**1**). If ARS is transferred along with a gene encoding leucine, approximately 5 to 20% of cells can replicate, but mitotic segregation is defective (**2**).

C. Centromeric sequences

CEN sequences are required for correct distribution of replicated chromosomes during mitosis. This has been shown by adding yeast CEN sequences to the plasmid in addition to the *Leu* gene and ARS. In this case, nearly 90% of progeny can grow on Leu^- medium (**1**) because normal mitotic segregation takes place (**2**). Thus, CEN sequences are necessary for normal distribution of the chromosomes at mitosis (see p. 92).

D. Telomeric sequences

When Leu^- yeast cells are transfected with a linear plasmid containing a Leu gene, and ARS and CEN sequences, they still fail to grow (**1**). The reason is that the linear plasmid does not contain TEL sequences. However, if TEL sequences are attached to both ends of the plasmid (**3**) before it is incorporated into the cells (**4**), normal replication and mitotic segregation take place (**4**). In this case, the linearized plasmid behaves as a normal chromosome. (Figures in B, C, D adapted from Lodsih et al, 2016).

Medical relevance

Losses or rearrangements of subtelomeric sequences are important causes of human developmental disturbances and are found in approximately 5% of patients with mental retardation (De Vries et al., 2003; Knight and Flint., 2000).

Further Reading

Albert B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Clarke L, et al. The structure and function of yeast centromeres. Annu Rev Genet 1985; 19: 29–55

De Vries BB, et al. Telomeres: a diagnosis at the end of the chromosomes. J Med Genet 2003; 40(6): 385–398

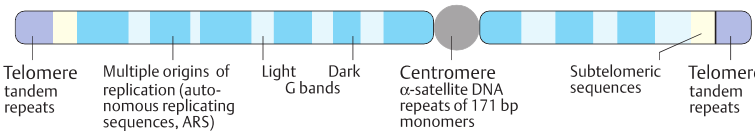
Kennard ML. Engineered mammalian chromosomes in cellular protein production: future prospects. Methods Mol Biol 2011; 738: 217–238

Knight SJ, Flint J. Perfect endings: a review of subtelomeric probes and their use in clinical diagnosis. J Med Genet 2000; 37(6): 401–409

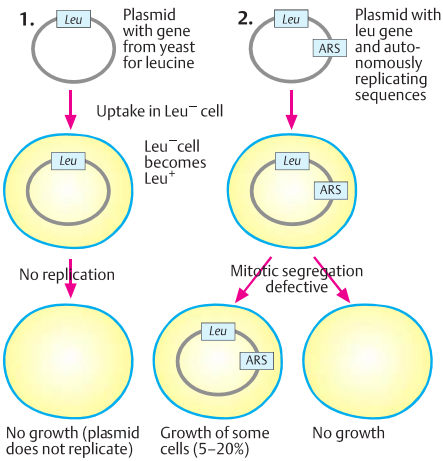
Lodish H, et al. Molecular Cell Biology. 6th ed. New York, NY: W. H. Freeman, 2016

Murray AW, Szostak JW. Construction of artificial chromosomes in yeast. Nature 1983; 305(5931): 189–193

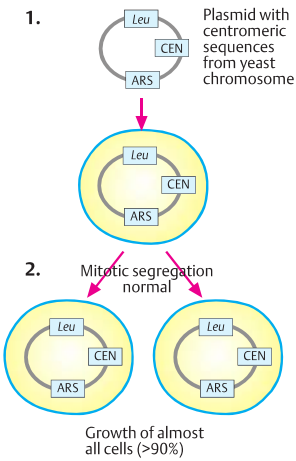
Schueler MG, et al. Genomic and genetic definition of a functional human centromere. Science 2001; 294 (5540): 109–115



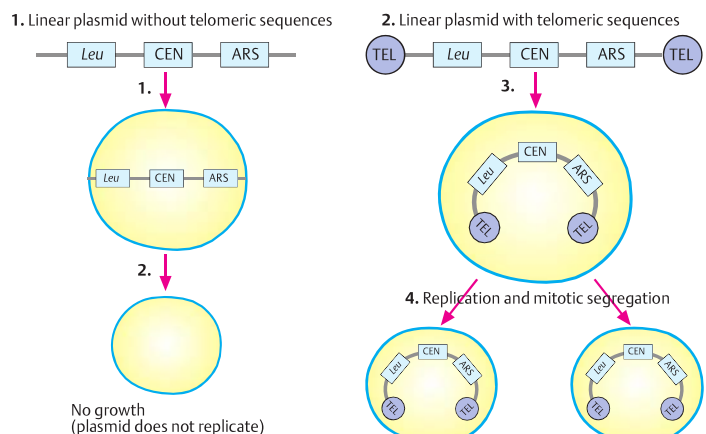
A. Basic features of a eukaryotic chromosome



B. Requirement for autonomous replicating sequences (ARS)



C. Requirement for centromeric sequences



D. Requirement for telomeric sequences

Nucleosomes

Nucleosomes are the basic subunit of eukaryotic chromosomes. They were first delineated by R. D. Kornberg and by Olins and Olins in 1974. A nucleosome consists of DNA-binding proteins that belong to two classes: histones and nonhistone proteins. DNA winds around a nucleosome in a defined pattern. The DNA double helix wound around each nucleosome allows a tight packing with a DNA packing ratio of approximately 1:10,000.

A. The nucleosome, the basic unit of DNA packing

A nucleosome consists of eight core histone molecules (octamer): two copies each of H2A, H2B, H3, and H4 (**1**). Around the disk-shaped octamer of histones are 140–150 base pairs of DNA (147 bp in humans) wrapped around 1.7 times. Linker DNA of variable length between 8 and 114 bp links two adjacent nucleosomes with each other. It is associated with histone H1.

Each nucleosome is separated from the other by 50 to 70 bp linker DNA, which yields a repeat length of 157 to 240 bp. The four histones are small proteins of 102 to 135 amino acids. More than one-fifth of the amino acids of the core histones are either lysine or arginine, two amino acids with basic side chains. Their positive charges can neutralize the negatively charged backbone of DNA. Each core histone has an N-terminal amino acid tail extending out of the DNA-histone core. For transcription and repair, the tight association of histones and DNA has to be loosened (see p. 180). H4 and H3 belong to the most conserved proteins in evolution; H2A and H2B are present in all eukaryotes, but their sequence varies between species.

B. Three-dimensional structure

The three-dimensional structure of the nucleosome (shown from above), based an X-ray diffraction structure at a high resolution of 2.8 Å, shows DNA wrapped around its histone core. One strand of DNA is shown in green, the other in brown. The histones are shown in different colors in the middle.

C. Chromatin structures

Chromatin occurs in different stages of compaction: condensed (tightly folded), less condensed (partially folded), and an extended, unfolded form. When extracted from cell nuclei in isotonic buffers, chromatin appears as fibers of approximately 30 nm in diameter. The corresponding electron microscopy photographs on the right, obtained by different techniques, show the condensed (folded) chromatin as compact 300 to 500 Å structures (top), as a 250-Å fiber when partially folded (middle), and as a "beads-on-a-string" 100-Å chromatin fiber (bottom). (Electronmicrophotograph from Thoma et al, 1979).

D. Chromatin segments

Chromatin occurs in segments of variable degrees of compaction. This corresponds to the third level of organization, the packing of the 30-nm fiber. This yields an overall packing ratio of 1,000 in euchromatin (about the same in mitotic chromosomes) and 10,000-fold in heterochromatin in both interphase and mitosis. (Image in B from Luger et al, 1997, with permission from Dr. TJ Richmond. Figure in D adapted rom Alberts et al, 2015).

Medical relevance

Faulty changes in chromatin structure cause different disorders and some forms of cancer (see Part III, p. 238, 312).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

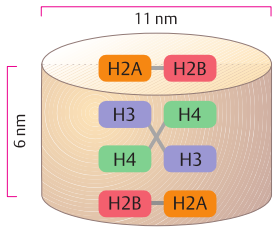
Luger K, et al. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 1997; 389(6648): 251–260

Olins AL, Olins DE. Spheroid chromatin units (v bodies). Science 1974; 183(4122): 330–332

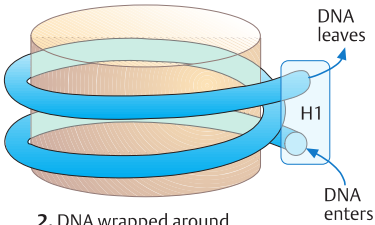
Richmond TJ Davey CA. The structure of DNA in the nucleosome core. Nature 2003; 423(6936): 145–150

Schalch T, et al. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature 2005; 436(7047): 138–141

Thoma F, et al. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol 1979; 83(2 Pt 1): 403–427



1. Nucleosome and histones

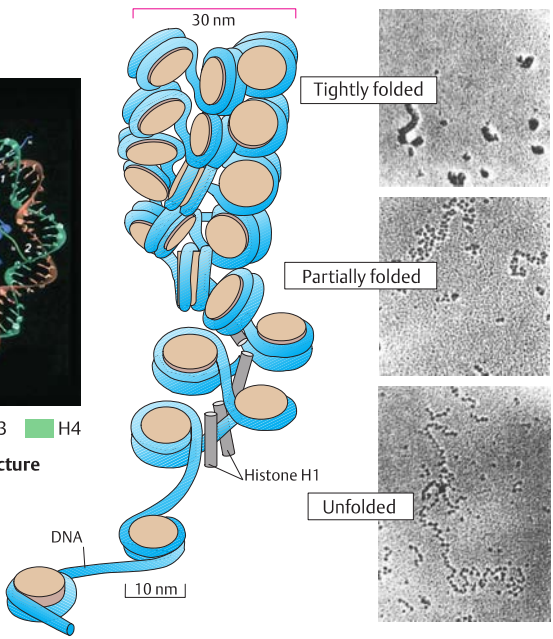


2. DNA wrapped around a nucleosome (140–150 bp depending on the species)

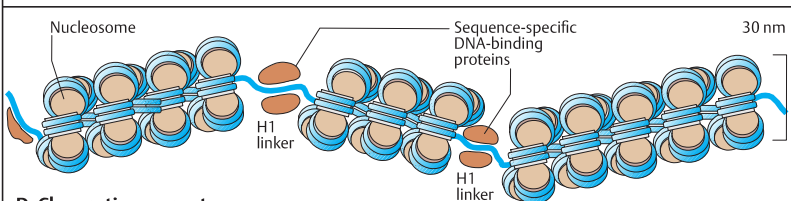
A. The nucleosome, the basic unit of DNA packing

H2A H2B H3 H4

B. Three-dimensional structure of a nucleosome



C. Chromatin structures



D. Chromatin segments

Packing DNA in Chromosomes

The total length of DNA in a eukaryotic cell is approximately 1 m. During mitosis, this needs to be packed into chromosomes of approximately 3 to 7 µm in length. This is equivalent to packing 401 km of London Underground into a suitcase (Kakui & Uhlmann, 2017). The packing occurs in a highly organized manner. The change from interphase to mitotic chromosomes requires a class of proteins called *condensins*. These use energy from adenosine triphosphate (ATP) hydrolysis to coil each interphase chromosome into a mitotic chromosome. Two of the five subunits of condensins interact with ATP and DNA. The cell cycle protein cdc2 (see Cell Cycle Control, p. 90) is required for both interphase and mitotic condensation.

A. Model for packing DNA in chromatin

Chromosomal DNA is folded and packed in six successive levels of hierarchical organization in a metaphase chromosome. These levels are shown schematically from top to bottom. First, a condensed section loops out of a metaphase chromosome (1,400 nm). Second, a higher magnification of this section shows a condensed section of a metaphase chromosome (at 700 nm). Third, DNA loops out from a chromosome scaffold at 300 nm resolution. Fourth, each loop corresponds to the 30-nm chromatin fiber of tightly packed nucleosomes shown at the next level below. Fifth, 11-nm "beads-on–string form" Sixth, the DNA double helix represents the ultimate resolution. (Figure adapted from Alberts et al, 2015 and Lodish, 2016).

B. Chromosome territories during interphase

Individual chromosomes occupy particular territories in an interphase nucleus. T. Cremer and coworkers (Bolzer et al., 2005) have shown that small chromosomes are located preferentially toward the center of fibroblast nuclei, whereas large chromosomes are positioned preferentially toward the nuclear rim. Measurements along the optical axes of the chromosome territories of human chromosomes 18 and 19 suggest that the gene-poor chromosome 18 is closer to the top or bottom of the nuclear envelope than chromosome 19. Probably, complex genetic and epigenetic mechanisms act at different levels to establish, maintain, or alter higher-order chromatin arrangements as required for proper nuclear functions. The numbers in (**1**) on the left indicate different degrees of chromosome decondensation (Monto Carlo relaxation steps 200, 1,000, and 400,000). (Images from Bolzer et al, 2005, provided by Thomas Cremer, München, Germany).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Bolzer A, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol 2005; 3(5): e157

Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2001; 2(4): 292–301

Cremer T, Cremer C. Chromosome territories. Cold Spring Harb Perspect Biol 2010; 2 (a003889)

Gilbert N, et al. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell 2004; 118(5): 555–566

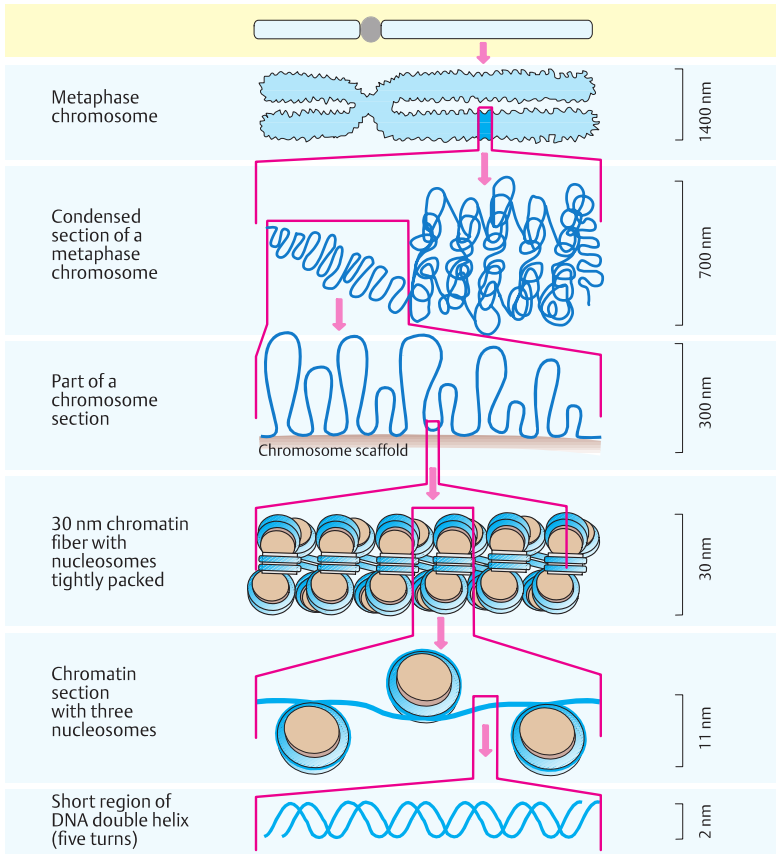
Hagstrom KA, Meyer BJ. Condensin and cohesin: more than chromosome compactor and glue. Nat Rev Genet 2003; 4(7): 520–534

Kakui Y, Uhlmann F. Building chromsomes without bricks. Science 2017; 356: 1233–1234

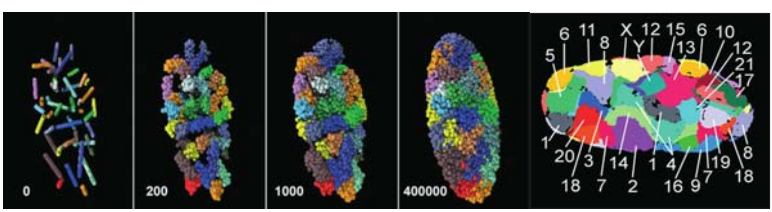
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Sun HB, et al. Size-dependent positioning of human chromosomes in interphase nuclei. Biophys J 2000; 79(1): 184–190

Tyler-Smith C, Willard HF. Mammalian chromosome structure. Curr Opin Genet Dev 1993; 3(3): 390–397



A. Model for packing of DNA in chromatin



1. Chromosome position in metaphase and interphase

2. Territories for each chromosome at interphase

B. Chromosome territories during interphase

The Telomere

The telomere at each end of a linear eukaryotic chromosome contains specialized repetitive DNA nucleotide sequences. These comprise approximately 2,500 nucleotide repeats (TTAGGG). The telomere protects its end of the chromosome from joining with other chromosomes. In addition, it helps overcome the problem a chromosome has with its ends at replication: since DNA synthesis occurs solely in the 5′ to 3′ direction, only the 3′ to 5′ template can be used continuously. That is, at the end of the 5′ to 3′ template (lagging strand), DNA cannot be synthesized because the final RNA primer cannot be attached. The unsynthesized bit of single-stranded DNA at the 3′ end would in effect be lost at each cell division. A large protein-RNA complex in the telomere, the *telomerase*, circumvents this problem. The name telomere is derived from the Greek words telos (τέλος – end) meros (μέρος – root).

A. General structure of the telomere

The telomeres in metaphase chromosomes can be specifically visualized by in situ hybridization techniques (**1**). In the terminal 6 to 10 kb of a chromosome, telomere and telomere-associated sequences can be distinguished from areas containing genes (**2**). DNA sequences would form a single-stranded end. A complex enzyme system, telomerase, prevents this. (Image obtained on 29 August 2016 at www.ucsf.edu/).

B. Replication by telomerase

Sequence-specific DNA-binding proteins attract an enzyme, telomerase. It binds to the incomplete 3′ end of the single-stranded telomeric DNA (**1**) and elongates it in the 5′ to 3′ direction by DNA synthesis using telomerase RNA as a template (**2**). After the parental strand has been extended, the lagging strand is completed by conventional DNA polymerases in the 5′ to 3′ direction (**3**).

C. Telomerase structure

The RNA of telomerase resembles other reverse transcriptases and can synthesize DNA using the RNA it contains as a template. Numerous telomerase-associated proteins (“fingers”) stabilize the region where telomerase RNA served as template for DNA synthesis (see Disorders Resulting from Defects in Telomeres, p. 242). At the end of a telomere the DNA forms a loop (t-loop). Its 3′ end displaces the 5′ end and inserts itself into the double-stranded telomeric DNA (not shown). Telomerase function is maintained in stem cells with high rates of proliferation. In contrast, in most other cells types the telomerase activity is reduced or absent, and telomeric DNA is lost at every cell division. As a result telomeres shorten over time. This limits the number of times a cell can divide. The cell enters a nondividing state called *replicative cell senescence*. In addition the protective protein caps deteriorate. (Figures in B and C adapted from Alberts et al, 2015).

Medical Relevance

See Disorders resulting from Defects in Telomeres (see p. 242).

Further Reading

Alberts B, et al., eds. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Armanios M, Blackburn EH. The telomere syndromes. Nature Rev Genet 2012; 13(10): 694–704

Blackburn EH et al. Human telomere biology: A contribution and interactive facor in aging, Science 2015; 350(6255): 1193–1198

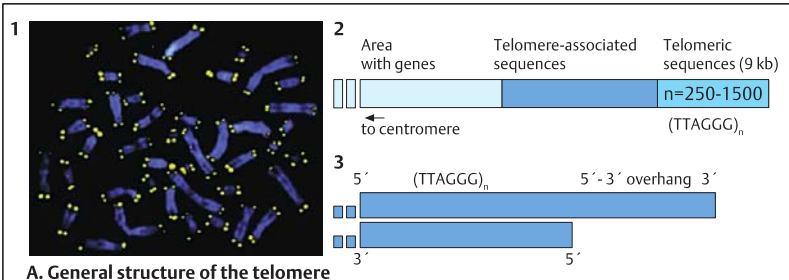
Brady C. Elizabeth Blackburn and the Story of the Telomeres. MIT Press, Cambridge, Massachusetts, 2007

Griffith JD et al. Mammalian telomeres end in a large duplex loop. Cell 1999; 97: 503–514

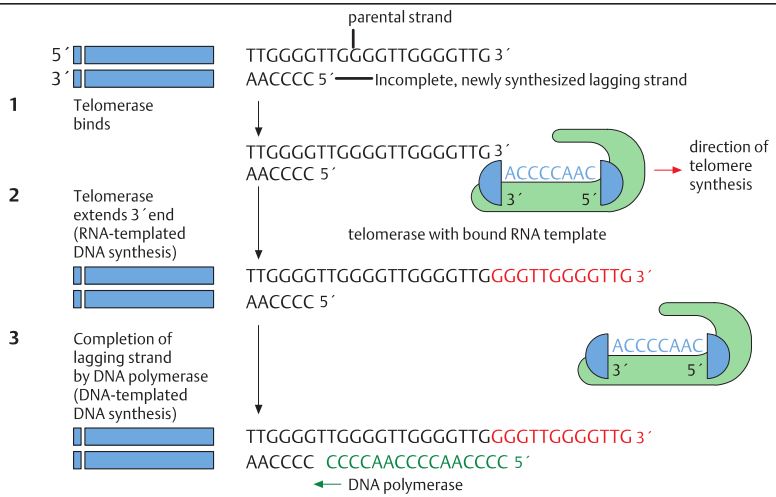
Holohan B et al. Telomeropathies: An emerging sprectrum disorder. J Cell Biol 2014; 205(3): 289–299

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2014

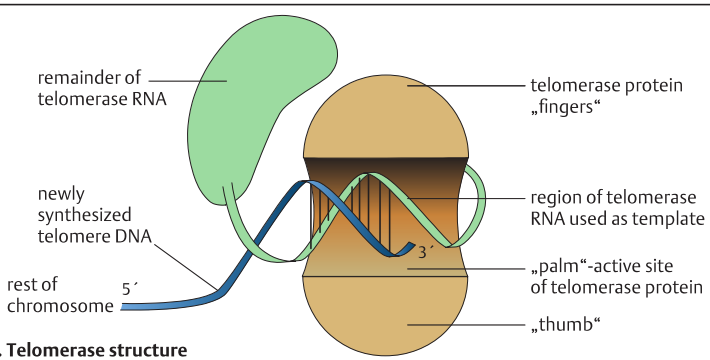
Savage SA, et al. The genetics and clinical manifestations of telomere biology disorders. Genet Med 2010; 12(12): 753–764



A. General structure of the telomere



B. Replication by telomerase



C. Telomerase structure

Chromosomes in Metaphase

In eukaryotes, the cellular genetic material (DNA and associated molecules) is packaged into chromosomes. Each chromosome contains a continuous thread of DNA double helix. The term *chromosome*, derived from the Greek word χρώμα (chrōma color) and σώμα (sōma, body), a term introduced by W. Waldeyer in 1888. In prokaryotes, the single chromosome usually has a circular structure. In eukaryotes, the chromosomes are located in the nucleus at defined positions. As first noted by W. Flemming in 1879, chromosomes can be visualized under a light microscope during mitosis as separate, individual structures. In 1956, Tjio and Levan in Lund, and Ford and Hamerton in Oxford established that man has 46 chromosomes, not 48, as had been previously assumed. Each organism has a defined number of chromosomes with distinct morphological appearances (the karyotype).

A. The metaphase chromosomes of man

A human metaphase is shown here at approximately 1,200-fold magnification. Each chromosome comes as a pair, one derived from the mother and one from the father. Each pair is defined by its length, position of the centromere, and the size and arrangement of transverse light and dark bands (banding pattern). Parts of a chromosome also can be identified by its banding pattern. In a typical metaphase preparation, approximately 400 to 550 distinct bands can be distinguished. In prometaphase, the chromosomes are longer than in metaphase and show more bands. Thus, for certain purposes, chromosomes are also studied in prometaphase. Different banding techniques exist for demonstrating selected features.

B. Types of metaphase chromosome

Each chromosome is classified as *submetacentric, metacentric,* or *acrocentric* according to the location of its centromere. This appears as a constriction and is the point of attachment of the spindle during mitosis. The centromere divides a submetacentric chromosome into a short arm (p arm, derived from the French word *petit*) and a long arm (q, next letter after p). In metacentric chromosomes, the short and long arms are about the same length. The short arm (p) of acrocentric chromosomes consists of a gene-free dense appendage called a *satellite* at the end of a stalk (not to be confused with satellite DNA).

C. Karyogram

A karyogram is a systematic visual representation of the karyotype. It displays all chromosomes in homologous pairs, arranged according to their relative lengths and the positions of their centromeres. Chromosomes are numbered as 22 pairs of chromosomes (autosomes, pairs 1–22) in addition to either two X chromosomes, in females, or an X and a Y chromosome, in males (karyotype 46,XX or 46, XY, respectively). The 22 pairs of autosomes in man are divided into seven groups (A–G). The karyotypes of different species of animals show differences and similarities depending on their evolutionary relationships.

Medical relevance

The analysis of chromosomes in metaphase (or under certain circumstances in prometaphase) is an important tool of diagnostics because approximately 1 in 400 newborns carries an aberration of the normal number or structure of a chromosome.

Further Reading

Caspersson T, et al. Differential binding of alkylating fluorochromes in human chromosomes. Exp Cell Res 1970; 60(3): 315–319

Dutrillaux B. Lejeune J. Sur une nouvelle technique d'analyse du caryotype humain. CR Acad Sci Paris D 1971; 272: 2638–2640

Ford CE, Hamerton JL. The chromosomes of man. Nature 1956; 178(4541): 1020–1023

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2014

McGowan-Jordan J, et al., eds. 2016: An International System for Human Cytogenetic Nomenclature. Basel, Switzerland: Karger, 2016

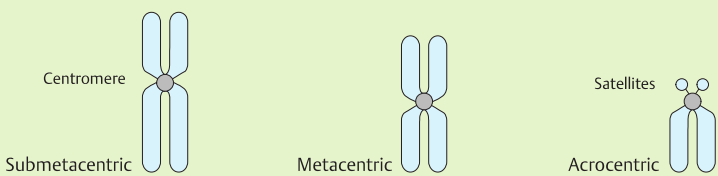
Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001

Riddihough G. Chromosomes through space and time. Science 2003; 301(5634): 779–802

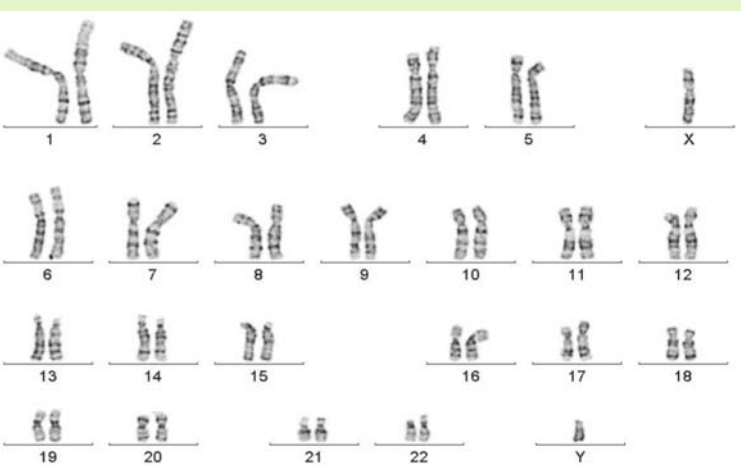
Speicher MR, et al. Vogel and Motulsky's Human Genetics. Problems and Approaches. Heidelberg: Springer, 2010

Tjio JH, Levan A. The chromosome number of man. Hereditas 1956; 42(1–2): 1–6

A. Microscopic appearance of metaphase chromosomes of man



B. Types of metaphase chromosome



C. Karyogram of the metaphase in A

The Banding Patterns of Human Chromosomes

With special staining techniques introduced in 1971, a reproducible, chromosome-specific pattern of transverse intensively and less intensively stained regions (bands) can be visualized along chromosomes in metaphase or prometaphase. This allows assigning structural alterations and gene loci to defined regions along a chromosome. When Tjio and Levan, and Ford and Hamerton, independently in 1956, determined the chromosome number of man to be 46, many of the chromosome pairs could not be distinguished. The complete set of chromosomes of an individual or a species is the *karyotype*. The chromosomes are arranged in homologous pairs in a *karyogram*. The most frequently used type of banding is *G-banding* (also referred to as GTG banding). This name is derived from the Giemsa stain, which is applied to the chromosomal preparation after it has been treated with the proteolytic enzyme trypsin. Other types of band are Q bands (quinacrine-induced), R bands (reverse of G), C bands (centromeric constitutive heterochromatin), T bands (telomeric), and others (see McGowan-Jordan et al., 2016 and Appendix, Table 1, p. 403).

A. Banding patterns and sizes of human chromosomes 1 to 12

A band is defined as part of a chromosome that can be clearly distinguished from its adjacent band and shows staining of different intensity. This plate schematically shows the G-banding pattern of human chromosomes 1 to 12 (for chromosomes 13–22, and the X and Y chromosomes, see next page). At standard resolution, approximately 450 individual bands can be distinguished per haploid set of human chromosomes. With variant techniques a higher resolution (800 bands) can be achieved.

Each chromosome is divided into regions along the short (p) arm and the long (q) arm. The regions are numbered from the centromere to the telomeres. For example, the proximal short arm of chromosome 1 begins (next to the centromere) with region 1, containing a dark and a light band, followed by regions 2 and 3. Within each region, bands are also numbered proximally to distally (toward the telomere). The number of each band is stated directly after the number of the region. For example, bands 2 and 3 of region 2 are designated 22 and 23, respectively; bands 1 to 6 of region 3 are designated 31, 32, 33, 34, 35, and 36, respectively. Each band is designated according to its chromosome, chromosome arm, region, and band. Thus, 1p23 indicates region 2, band 3 of the short arm of chromosome 1. Additional bands that become visible with higher resolution are indicated by a decimal: if one additional band can be distinguished in 1p21, this region is designated as 1p2, 1p21.2, 1p21.3, and so on. This system is outlined in detail in McGowan-Jordan et al. (2016). The size of each chromosome in million base pairs is given below each chromosome.

Medical relevance

The banding pattern is used to define the location of structural chromosomal aberrations and the location of genes as defined by light microscopy on a chromosome (for the map of gene loci relevant for mongenic disorders, see p. 392).

Further Reading

Ford CE, Hamerton JL. The chromosomes of man. Nature 1956; 178(4541): 1020–1023

Gardener RJM, et al. Chromosome Abnormalities and Genetic Counseling. 4th ed. Oxford: Oxford University Press, 2012

Gersen SL, Keagle MB, eds. The Principles of Clinical Cytogenetics. 2nd ed. Totowa, NJ: Humana Press, 2005

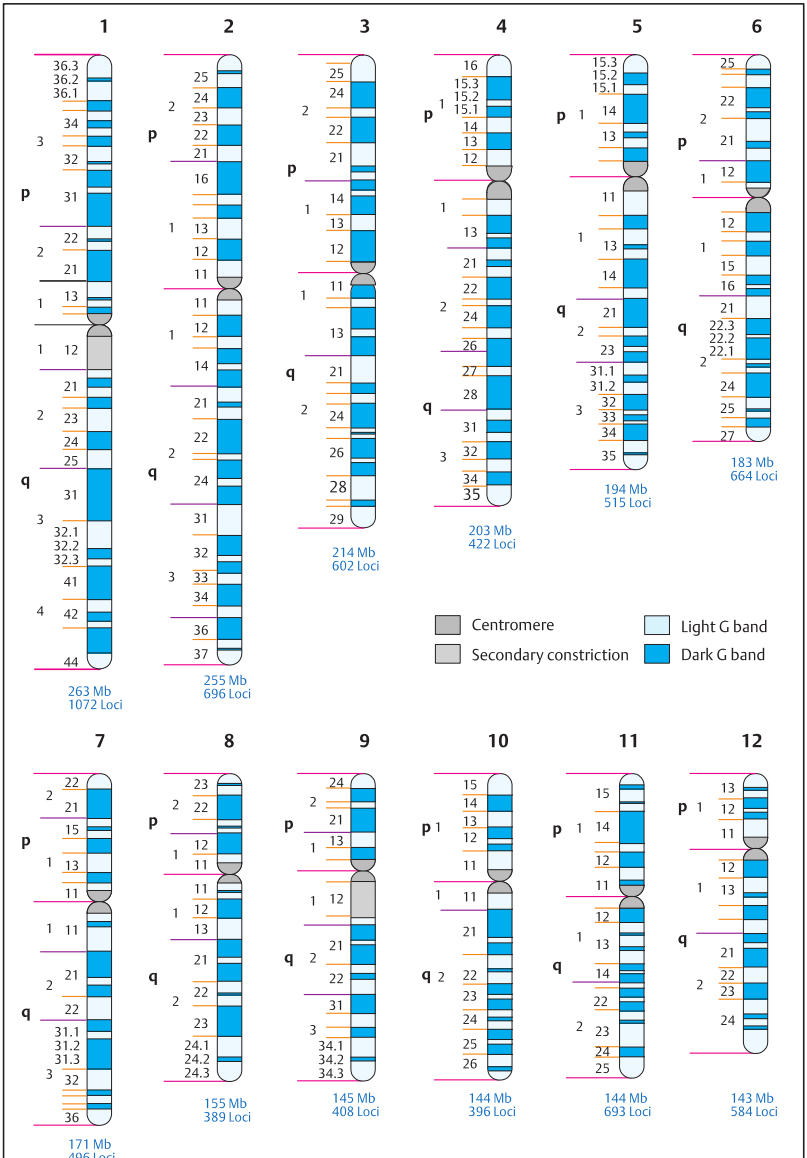
McGowan-Jordan J, et al., eds. ISCN 2016: An International System for Human Cytogenetic Nomenclature. Basel, Switzerland: Karger, 2016

Mikhail FM. Chromosomal Basis of Inheritance. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013 (e-book)

Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001

Seabright M. A rapid banding technique for human chromosomes. Lancet 1971; 2(7731): 971–972

Tjio JH, Levan A. The chromosome number of man. Hereditas 1956; 42(1–2): 1–6



A. Banding patterns and sizes of human chromosomes 1–12

Karyotype of Man and Mouse

The 22 chromosome pairs of man are arranged in seven groups (A–G): group A consists of three pairs, chromosomes 1 to 3, which are large submetacentric chromosomes; group B (chromosomes 4–5) consists of two pairs of fairly long submetacentric chromosomes, which are slightly smaller than group A; members of group C (chromosomes 6–12) are all submetacentric; members of group D (chromosomes 13–15) are all acrocentric; members of group E (chromosomes 16–18) are all submetacentric; members of group F (chromosomes 19–20) are all metacentric; and members of group G (chromosomes 21–22) are all small acrocentric chromosomes. In addition, the X and Y chromosomes do not belong to a group.

A. Banding patterns of human chromosomes 13 to 22, X, and Y

The schematic representation of the human karyotype is continued from the previous page.

B. Karyotype of the mouse

The mouse (*Mus musculus*) has 20 chromosome pairs, each with a specific banding pattern (**1**). All chromosomes except the X chromosome are acrocentric, with the centromere at the very end.

Variant strains of mice have a karyotype with metacentric chromosomes. These are the result of centric fusion between two acrocentric chromosomes (**2**). The example shows fused chromosome pairs 4 and 2 (4/2); 8 and 3; 7 and 6; 13 and 5; 12 and 10; 14 and 9; 18 and 11; and 17 and 16. Chromosome pairs 1, 15, and 19 are unchanged, as in the standard karyotype. The fused chromosomes are present in all mice of this particular population. They represent an example of chromosome evolution by fusion. (Photographs: 1 from Traut, 1991; 2 provided by Dr. H. Winking, Lübeck, Germany).

Human chromosome nomenclature

An elaborate system of nomenclature to describe also complex chromosome changes in a brief formula is described in the International System for Human Cytogenetic Nomenclature (ISCN 2016; McGowan-Jordan et al.). The most important examples are listed in the Appendix, Table 2 (see p. 404).

Examples of Nomenclature Usage:

46,XX, normal female karyotype; 46,XY, normal male; 47,XXY, an extra X chromosome in a male karyotype; 45,X, a missing X or Y chromosome; 47,XY,+21, an extra chromosome 21 in a male karyotype; 2q, material missing from the long arm of chromosome 2 (a + indicates extra material); t(2;5)(q21;q31), reciprocal translocation between chromosome 2 and 5 with breakpoints at 2q21 and 5q31; t(13q14q), a centric fusion of chromosome 13 and chromosome 14; del, a deletion; dup, a duplication; inv, an inversion; ins, insertion; ins(5;2)(p14;q21), insertion of region q21 of chromosome 2 into region p14 of chromosome 5; mat-, pat-, maternal/paternal origin; der, derivative chromosome in a structural change; I, isochromosome; r, ring-shaped chromosome; h, heterochromatin.

Further Reading

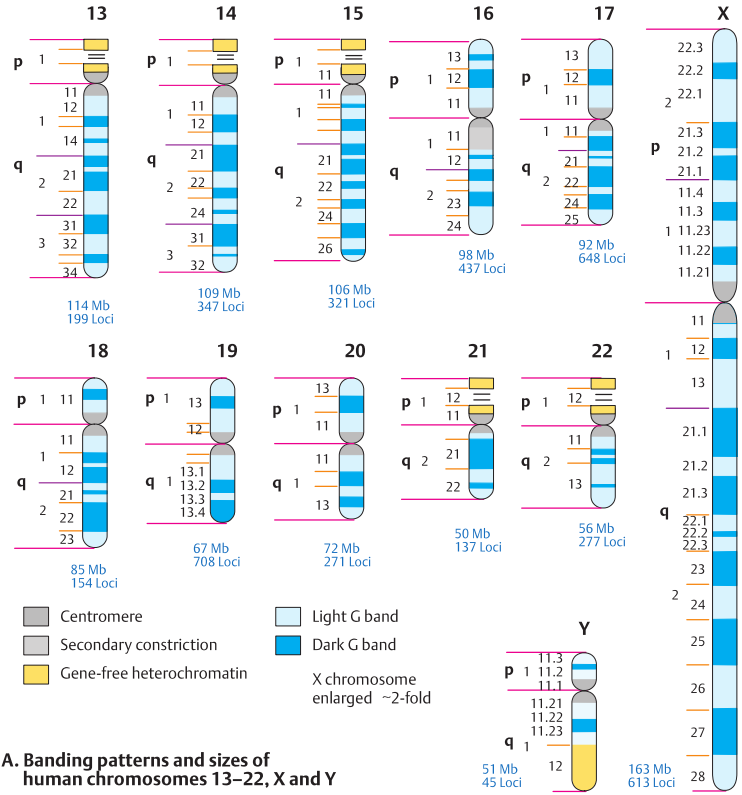
Gersen SL, Keagle MB, eds. The Principles of Clinical Cytogenetics. 2nd ed. Totowa, NJ: Humana Press, 2005

Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001

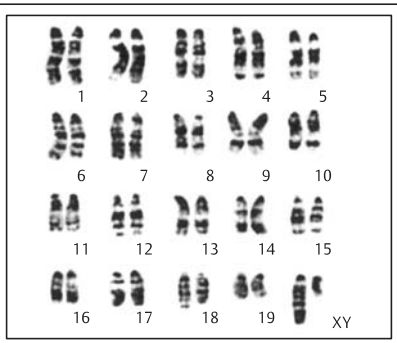
McGowan-Jordan J, et al., eds. ISCN 2016: An International System for Human Cytogenetic Nomenclature. Basel, Switzerland: Karger, 2016

Schwartz S, et al. Chromosome disorders. In: Longo DL, et al., eds. Harrison's Principles and practice of Internal Medicine. 18th ed. New York, NY: McGraw Hill, 2012: 509–518

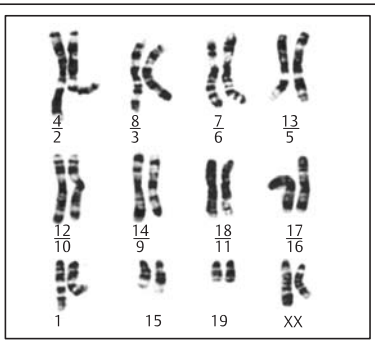
Traut W. Chromosomen. Klassische und molekulare Cytogenetik. Heidelberg: Springer-Verlag, 1991



A. Banding patterns and sizes of human chromosomes 13–22, X and Y



1. Standard



2. Variants in a population with fused chromosomes

B. Karyotype of the mouse (*Mus musculus*)

Preparation of Metaphase Chromosomes for Analysis

Chromosomes can only be analyzed when a cell is in metaphase (or occasionally prometaphase). These are usually obtained from peripheral blood lymphocyte cultures. For special purposes, metaphases can also be obtained from cultured skin fibroblasts, cultured amniocytes from amniotic fluid, and cultured chorionic cells from chorionic villi. From the bone marrow, cells in metaphase can be directly observed without culture. Peripheral blood lymphocytes grow in a suspension culture after having been stimulated to divide by phytohemagglutinin, a plant lectin that induces cell division in T-lymphocytes. The life span of lymphocytes is limited to a few cell divisions. However, by exposing the culture to Epstein–Barr virus, they can be transformed into a lymphoblastoid cell line with permanent growth potential (see p. 102).

A. Chromosome analysis from blood

Five principal steps are required: (i) lymphocyte culture, (ii) harvest of metaphase chromosomes, (iii) chromosome preparation, (iv) staining with special dyes that stain chromosomes (and chromatin), and (v) analysis by microscopy, which is now assisted by computer analysis.

Approximately 2 to 3 mL of whole blood with heparin added to approximately one-tenth of the blood volume is needed. Heparin must be added to prevent blood clotting. Phytohemagglutinin-stimulated lymphocyte culture for approximately 72 hours at 37°C yields metaphases after two cell divisions. Cells reaching mitosis are arrested in metaphase by adding a colchicine derivative (colcemid) for 2 hours prior to harvest. The culture is then terminated and the cells in metaphase are harvested (**1**).

At harvest, the culture solution is centrifuged (**2**). Hypotonic potassium chloride solution (0.075 M) is added for 20 minutes (**3**), after which a fixative solution (3:1 mixture of methyl alcohol and glacial acetic acid) is added (**4**). The fixed cells are taken up in a pipette, dropped onto a clean, wet, fat-free glass slide suitable for microscopic analysis, and air-dried (**5**). The preparation is treated according to the type of bands desired (**6**), stained (**7**), and the slide is covered with a cover glass (**8**). Suitable metaphases are located under the microscope at approximately 100-fold magnification and are subsequently examined at approximately 1,250-fold magnification (**9**). More than one cell has to be analyzed. Depending on the purpose of the analysis, between 5 and 100 metaphases (usually 10–20) are examined. Some of the metaphases are photographed under the microscope and subsequently can be cut out from the photograph (karyotyping). In this way, a karyogram can be obtained from the photograph of a metaphase (**10**). The time needed for a chromosome analysis varies depending on the problem, but it is usually 3 to 4 hours. Analysis and karyotyping time can be considerably shortened by computer procedures. The karyogram can be obtained by computer-assisted analysis.

Additional methods for chromosomal analysis: a variety of methods exist that do not use metaphase chromosomes for analysis. Examples are interphase fluorescence in situ hybridization (FISH), comparative genomic hybridization (CGH; see p. 218), fiber FISH, chromosome sorting by flow cytometry, and others.

Medical relevance

Cytogenetics with chromosomal analysis is an important tool in the diagnosis of genetic disorders that can be diagnosed by light microscopy. Under given circumstances, metaphase analysis can be replaced by molecular techniques, for example, array CGH (see p. 218).

Further Reading

Arakaki DT, Sparkes RS. Microtechnique for culturing leukocytes from whole blood. Cytogenetics 1963; 2: 57–60

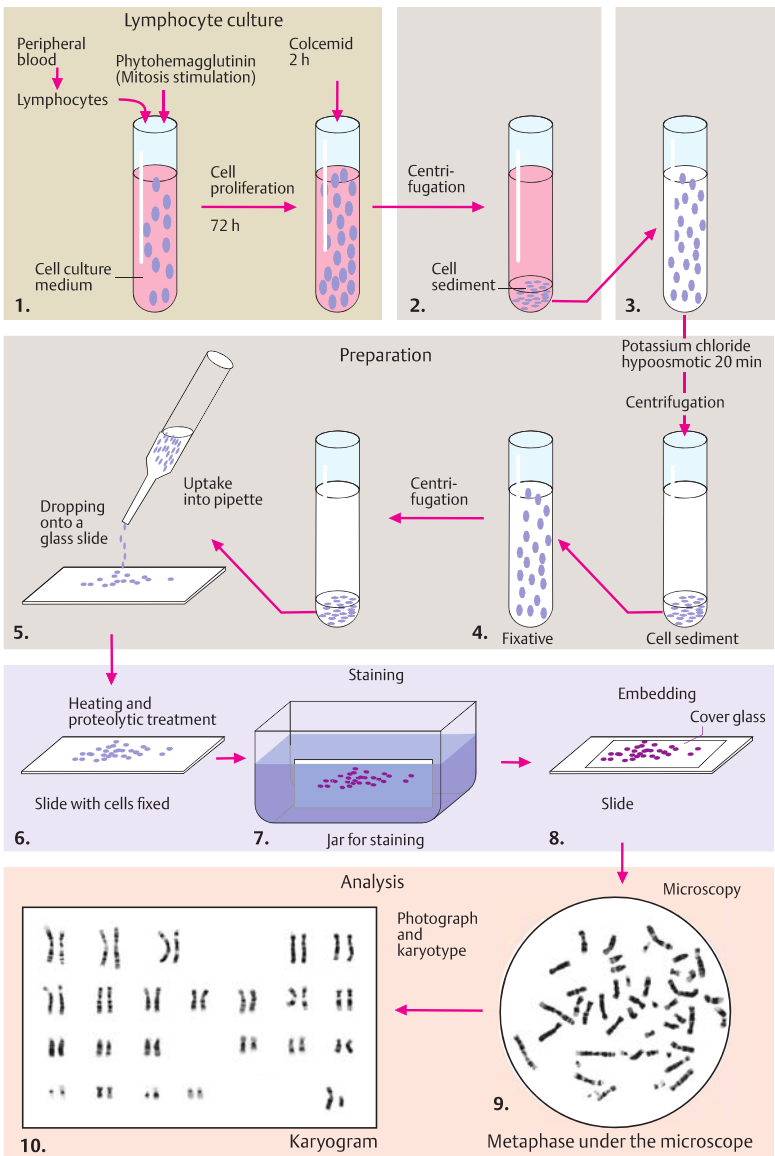
Gosden JR, ed. Chromosome Analysis Protocols. Methods in Molecular Biology. Vol. 29. Totowa, NJ: Humana Press, 1994

Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001

Moorhead PS, et al. Chromosome preparations of leukocytes cultured from human peripheral blood. Exp Cell Res 1960; 20: 613–616

Schwarzacher HG, et al., eds. Methods in Human Cytogenetics. Berlin: Springer-Verlag, 1974

Sumner T. Chromosomes. Organization and Function. Oxford: Blackwell, 2003



A. Chromosomal analysis from blood

Fluorescence *in Situ* Hybridization

Fluorescence *in situ* hybridization (FISH) analysis combines cytogenetic methods with DNA molecular analysis (molecular cytogenetics). DNA is denatured (made single-stranded) and the resulting single-strand hybridized with a single-stranded probe that only hybridizes at complementary sequences. *In-situ* hybridization was introduced by Mary-Lou Pardue and Joseph Gall at Yale University in 1969. It has developed into a variety of molecular genetic techniques that are applied to chromosome preparations in metaphase or interphase nuclei, an approach called *molecular cytogenetics*. Conventional metaphase chromosomal analysis can detect the loss or gain of chromosomal material of approximately 4 million base pairs or more, whereas FISH analysis can detect much smaller rearrangements.

A. Principle of FISH

Metaphase chromosomes or interphase cells fixed on a glass slide (*in situ*) are denatured (**1a**) into single-stranded DNA (**2**). A DNA probe (**1b**) labeled to allow recognition of its site of hybridization to a specific site (**3**). This is visualized by fluorescence in dark-field microscopy by a fluorescent-dye-labeled antibody labeled with biotin (**4**). To this primary antibody, a secondary antibody (here, a biotinylated antiavidin antibody) is attached (**5**). The resulting signal is amplified by attaching additional labeled antibodies (**6**).

B. FISH analysis in metaphase

Two metaphase chromosomes 22 have hybridized to a probe specific for the centromeric region of chromosome 22 (green signals) and chromosomal region 22q11 (red signals). The two red signals indicate normal presence of this region in each chromosome. In case of a deletion, this signal would not be present and indicate the loss in the affected chromosome. In case of a duplication, there would be three red signals. All other metaphase chromosomes are stained dark blue by DAPI (4′,6-diamidino-2-phenylindole).

C. Interphase FISH analysis

An interphase nucleus shows two green and three red signals. The three red signals indicate the presence of three chromosomes 21 in trisomy 21 (probes specific for chromosome region 21q22.13). The green signal is due to cohybridization with chromosomal regions 13q14.2.

D. FISH analysis of a translocation

This shows a reciprocal translocation between the long arm of chromosome 8 and the short arm of chromosome 4. This yields two derivative chromosomes, der4 and der8, resulting from the translocation. The region embracing the breakpoint on 8 at q24 was labeled by a 170-kilobase yeast artificial chromosome. As a result, three signals are visible: one over the normal chromosome 8, one over the altered chromosome 8 (der8), and one over the altered chromosome 4 (der4). Centromere-specific probes identify the four chromosomes involved. (Photograph courtesy of HJ Lüdecke, Essen.)

E. FISH of a complex rearrangement

In this partial metaphase, the short arm of chromosomes 5 (5p) is stained in green and the long arm (5q) in orange. The short arm of chromosomes 11 (11p) is stained in blue and the long arm (11q) in red. A reciprocal translocation between chromosome 5 and chromosome 11 is indicated by a derivative chromosome 11 (T1) and a derivative chromosome 5 (T2), which in addition has a pericentric inversion. The probes used hybridize specifically to the chromosome arms involved. The other chromosomes are counterstained by DAPI. (Photographs in B: courtesy A. Scheduikat & G. Siegert, Leipzig; in C: A.K. Teichmann, Leipzig; in D: H. J. Lüdecke, Düsseldorf, formerly Essen, and in E: G. Schwanitz, Bonn).

Further Reading

Cassidy A, Jones J. Developments in *in situ* hybridisation. Methods 2014; 70(1): 39–45

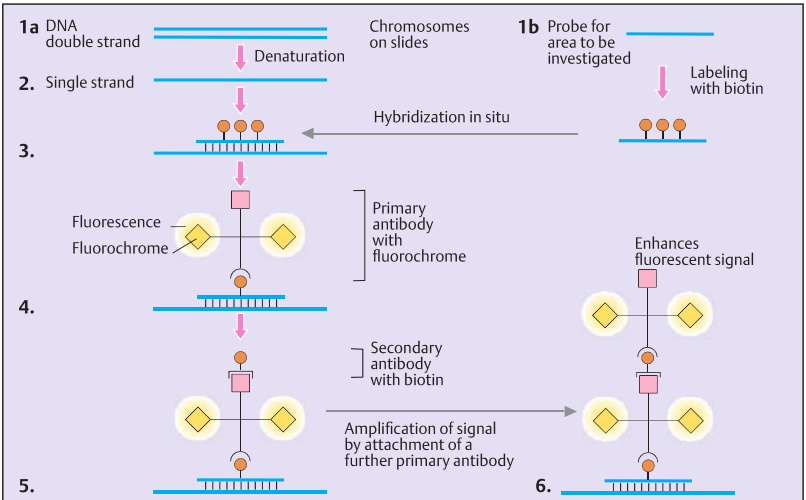
Jensen E. Technical review: *in situ* hybridization. Anat Rec (Hoboken) 2014; 297(8): 1349–1353

Liehr T, et al. The current state of molecular cytogenetics in cancer diagnosis. Expert Rev Mol Diagn 2015; 15(4): 517–526

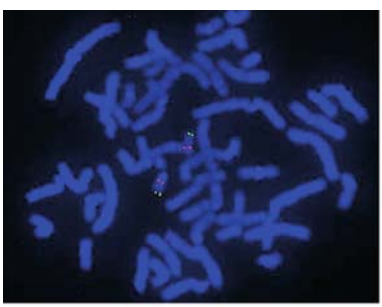
Pinkel D, et al. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci U S A 1986; 83(9): 2934–2938

Ried T, et al. Chromosome painting: a useful art. Hum Mol Genet 1998; 7(10): 1619–1626

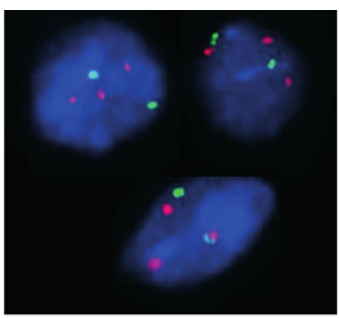
Speicher MR, Carter NP. The new cytogenetics: blurring the boundaries with molecular biology. Nat Rev Genet 2005; 6(10): 782–792



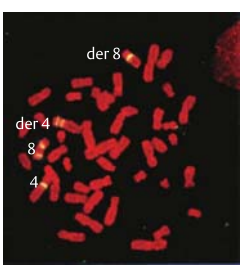
A. Principle of fluorescence-*in-situ*-hybridization



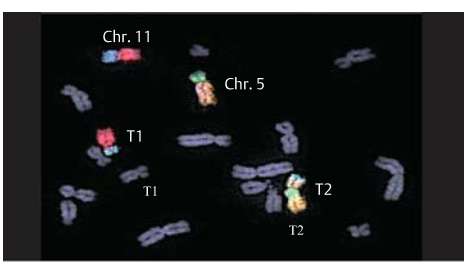
B. FISH analysis in metaphase



C. FISH analysis in interphase



D. Translocation 4;8



E. FISH analysis of a complex rearrangement

Multicolor Fluorescence *In Situ* Hybridization (FISH) Identification of Chromosomes

Computer-based methods can identify each chromosome pair individually and detect small rearrangements that escape conventional metaphase analysis. When sets of chromosome-specific probes hybridize to an entire chromosome or parts of a chromosome, the result is referred to as *chromosome painting*. To increase the number of chromosomes that can be differentiated at one time, a combination of labels (DNA probes that are labeled with more than one type of fluorochrome) and different ratios of labels are applied. The resulting mixed colors are detectable by automated digital image analysis. Two approaches have proved particularly useful: multiplex fluorescence in situ hybridization (FISH; Speicher et al., 1996) and spectral karyotyping (Schröck et al., 1996). Other approaches and modifications exist, for example, the use of artificially extended DNA or chromatin fibers (fiber FISH).

A. Multicolor FISH

Multicolor FISH (Speicher et al., 1996) uses sets of chromosome-specific DNA probes, which are hybridized to denatured metaphase chromosomes. Each probe is contained in a yeast artificial chromosome and labeled with a combination of DNA-binding fluorescent dyes specific for each chromosome. Five different fluorophores can produce 24 different colors for image analysis by epifluorescence microscopy using a charge-coupled device (CCD) camera. An image of each chromosome in a pseudocolor is displayed by appropriate computer software.

B. Spectral karyotyping

Spectral karyotyping (SKY; Schröck et al., 1996) combines optical microscopy, Fourier spectroscopy, and CCD imaging.

First, all 24 human chromosomes (chromosomes 1–22, X, and Y) are separated by flow sorting and labeled with at least one, but possibly as many as five fluorochrome combinations (SKY probes). Following DNA denaturation, the SKY probe is hybridized to a denatured (single-stranded) metaphase chromosome preparation at 37°C for 24 to 72 hours.

An interferometer and a CCD camera (SpectraCube) are coupled to an epifluorescence microscope to visualize the SKY probes. The emitted light is sent through a Sagnac interferometer and focused onto the CCD camera. In this way, an interferogram based on the optical path difference of the divided light beam is measured for each pixel of the image. The resulting unique spectral "signature" of each chromosome based on the raw spectral information is displayed as an RGB (red, green, blue) image of the metaphase and is used for automated chromosome classification. During this process, a discrete false color is assigned to all pixels of the image with identical spectra, revealing structural and numerical chromosome aberrations. Spectral karyotyping has a wide range of diagnostic applications in clinical and cancer cytogenetics.

Medical relevance

Computer-assisted cytogenetic analysis is widely used to identify structural chromosomal aberrations below the resolution power of light microscopy.

Further Reading

Chrombios. Molecular Cytogenetics. Multicolor FISH. Available at http://www.chrombios.com/cms/website.php. Accessed November 24, 2016

Imataka G, et al. Chromosome analysis using spectral karyotyping (SKY). Cell Biochem Biophys 2012; 62 (1): 13–17

Liehr T, ed. Fluorescence in Situ Hybridization (FISH). Heidelberg: Springer-Verlag, 2010

Liehr T, et al. Multicolor FISH probe sets and their applications. Histol Histopathol 2004; 19(1): 229–237

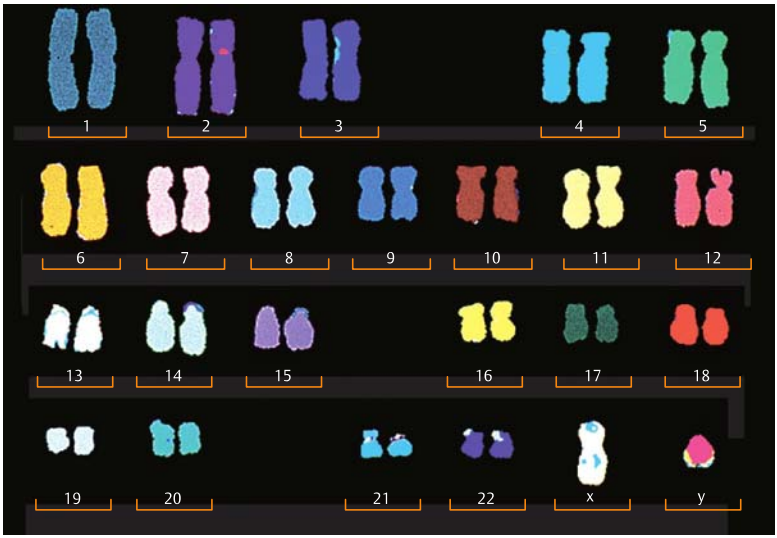
Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001

Ried T, et al. Chromosome painting: a useful art. Hum Mol Genet 1998; 7(10): 1619–1626

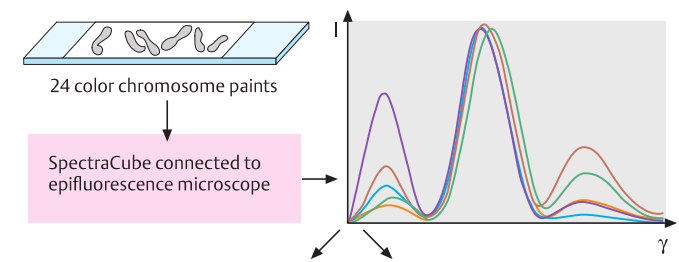
Schröck E, et al. Multicolor spectral karyotyping of human chromosomes. Science 1996; 273(5274): 494–497

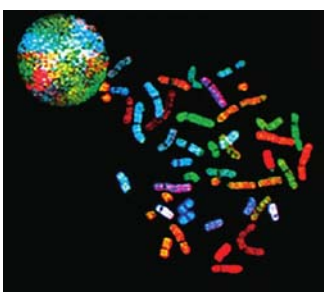
Speicher MR, et al. Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat Genet 1996; 12(4): 368–375

Uhrig S, et al. Multiplex-FISH for pre- and postnatal diagnostic applications. Am J Hum Genet 1999; 65 (2): 448–462

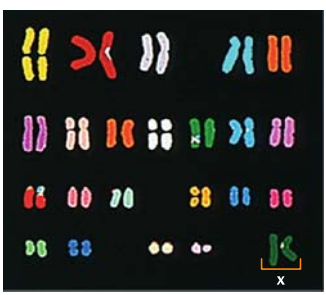


A . Multicolor fluorescence in–situ hybridization karyogram





Display colors in metaphase



Classification colors in the karyogram

B. Spectral karyotyping

Aneuploidy

Aneuploidy, a term introduced by Täckholm in 1922, is a deviation of the normal chromosome number. It leads to loss or gain of one or several individual chromosomes from the diploid set. A loss is called *monosomy*, a gain is called *trisomy*. Both result from nondisjunction at meiosis (see p. 94). Nondisjunction was discovered 1913 in *Drosophila melanogaster* by C. B. Bridges, who coined the term.

A. Trisomy in *Datura stramonium*

A. F. Blakeslee (1874–1954) in 1922 described phenotypic effects of trisomy in a plant, *Datura stramonium*, the common thorn apple (Jimson weed, also called *Devil's snare*). Capsules of plants with three copies of one of the 12 chromosomes had a characteristic appearance (phenotype). This suggested that each of the 12 chromosomes of this plant had its own genetic content that differed from all others. (Figure adapted from Blakeslee, 1922).

B. Trisomy in the mouse

Specific phenotypic effects of different autosomal trisomies can also be observed in mice. In the 1970s, A. Gropp and coworkers determined a specific developmental profile of morphological changes and malformations for each trisomy for one of the 19 pairs of autosomes (**1**). Embryos with a monosomy died very early, within the first 8 days of the 21-day gestation. The intrauterine survival time depended on the chromosome that was trisomic. For trisomy 2, 15, 3, and 7, the intrauterine survival time was the shortest (10–12 days); and for trisomy 13, 14, 16, 18, and 19, it was the longest (18–20 days). The phenotypes also varied, depending on the chromosome present in triplicate. In a mouse embryo with trisomy 12 (**2**) or trisomy 19 at birth (**3**), the brain showed different types and degrees of malformations compared with a normal control. The trisomic mice were generated by breeding mice with translocations among various chromosomes. Figures in 1: Gropp, 1982; 2 and 3 provided by Dr. H. Winking, Lübeck, Germany).

C. Nondisjunction in meiosis I or II

Nondisjunction may occur during either meiosis I or II. During meiosis I, one daughter cell will receive two chromosomes instead of one, whereas the other daughter cell receives none (in the opposite figure, only one pair of chromosomes is shown). As a result, gametes will be formed carrying either two chromosomes (disomy) or none (nullisomy), instead of one each. If nondisjunction occurs during meiosis II, the first meiotic division (see p. 96) is normal. However, during the second meiotic division, one daughter cell will receive two chromosomes (becoming disomic), and the other will receive none (nullisomic). After fertilization, a disomic gamete gives rise to a trisomic zygote, and a nullisomic gamete gives rise to a monosomic zygote.

D. Common types of numerical aberrations

Three states of aneuploidy can be distinguished: (i) trisomy (three chromosomes), (ii) monosomy (one chromosome instead of two), and (iii) triploidy and tetraploidy (all chromosomes present in triplicate or quadruplicate). Triploidy does not result from nondisjunction at meiosis, but from one of a variety of other processes. Two sperm cells may have penetrated the egg (dispermy), the egg or sperm may have an unreduced chromosome set as a result of restitution in the first or second meiotic division, or the second polar body may have reunited with the haploid egg nucleus. With dispermy, two of the three sets of chromosomes will be of paternal origin, resulting in 69,XYY, 69,XXY, or 69, XXX.

Further Reading

Avery AG, et al. Blakeslee: The Genus Datura. New York, NY: Ronald Press Co., 1959

Blakeslee AF. Variations in Datura due to changes in chromosome number. Am Nat 1922; 56(642): 16–31

Bridges CB. Non-disjunction of the sex chromosomes of Drosophila. J Exp Zool 1913; 15(4): 587–606

Boué A, et al. Cytogenetics of pregnancy wastage. Adv Hum Genet 1985; 14: 1–57

Epstein CJ. Down syndrome (trisomy 21). In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 1223–1256. Available at www.ommbid.com

Gropp A. Value of an animal model for trisomy. Virchows Arch A Pathol Anat Histol 1982; 395(2): 117–131

Täckholm G. Zytologische Studien über die Gattung *Rosa*. Act Hort Bergiani 1922; 7: 98–381

Normal

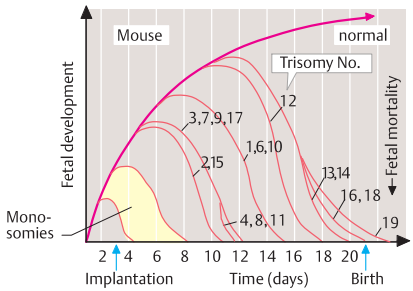
1 Rolled 2 Glossy

3 Buckling 4 Elongate

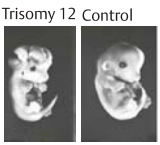
5 Echinus 6 Cocklebur 7 Microcarpic 8 Reduced

9 Poinsettia 10 Spinach 11 Globe 12 Ilex

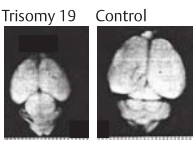
A. Trisomy in jimsonweed (*Datura stramonium*)



1. Developmental course of trisomies

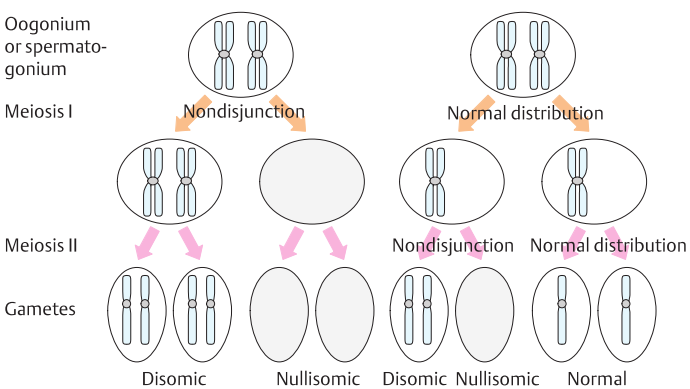


2. Mouse embryo

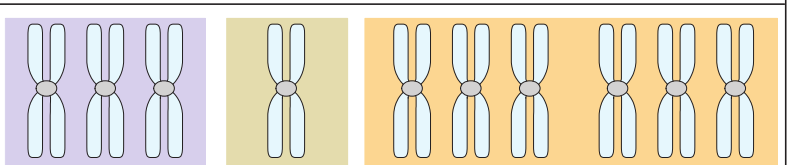


3. Brain

B. Trisomies in the mouse



C. Nondisjunction in meiosis I or meiosis II



1. Trisomy **2.** Monosomy **3.** Triploidy (all chromosomes in triplicate)

D. Common types of aneuploidy

Chromosome Translocation

In cytogenetics, translocation refers to a structural change involving a break in the chromosomes involved. It results in a change in position of a part of (or a whole) chromosome from its usual location to another, without changing the total number of chromosomes or genes. Usually, a translocation is a reciprocal event, one part being exchanged with another. A special type of translocation, the fusion of two acrocentric chromosomes, is called a *Robertsonian translocation*, after W. R. B. Robertson who first observed it in insects in 1911.

A. Reciprocal translocation

A reciprocal translocation between two chromosomes is called *balanced* when it results in loss or gain of part of a chromosome.

During meiosis, the normal homologous chromosomes and the chromosomes are involved in the reciprocal translocation pair. Each of the chromosomes pairs with its homologous partner. However, the translocation chromosomes cannot pair directly but form a quadriradial configuration. The distribution of the derivative chromosome can result in three outcomes: (i) alternate, (ii) adjacent-1, and (iii) adjacent-2. Alternate segregation results in chromosomally balanced gametes. Adjacent-1 and adjacent-2 are unbalanced.

In *adjacent-1* segregation, the gametes receive a normal chromosome and a translocation chromosome. The translocation chromosomes are unbalanced because parts are missing or added.

In the rare *adjacent-2* segregation, homologous centromeres go to the same pole. This produces two generally more extreme types of unbalanced gamete (1:3 and 0:4 segregation).

B. Centric fusion of acrocentric chromosomes

Robertsonian translocations (centric fusion) involve either a pair of homologous or two nonhomologous acrocentric chromosomes. In homologous fusion during meiosis, only disomic and nullisomic gametes are produced. Fusion of nonhomologous acrocentric chromosomes is much more common. Chromosome 14 and chromosome 21 (**1**) are most frequently involved in centric fusion (in ~1 in 1,000 newborns). When the long arm of chromosome 21 (21q) and the long arm of chromosome 14 (14q) fuse to form chromosome 14q21q (**2**), four types of gamete can be formed: normal, balanced, disomic, or nullisomic for chromosome 21 (**3**). After fertilization, the corresponding zygotes contain only one chromosome 21 (nonviable monosomy 21), a normal chromosome complement, a balanced chromosome complement with the fused chromosome, or three chromosomes 21 (trisomy 21).

Medical relevance

Chromosomal translocations are a common cause of chromosomal imbalance in offspring and an increased risk familial occurrence in siblings. Some translocations are more common than expected by chance. Nearly 90% of translocations with a breakpoint in chromosome 11q23 also involve chromosome 22q11. A similarity of DNA sequence resulting from genomic duplication events may predispose to such aberrations (see Part II, Genomic Disorders, p. 236).

Further Reading

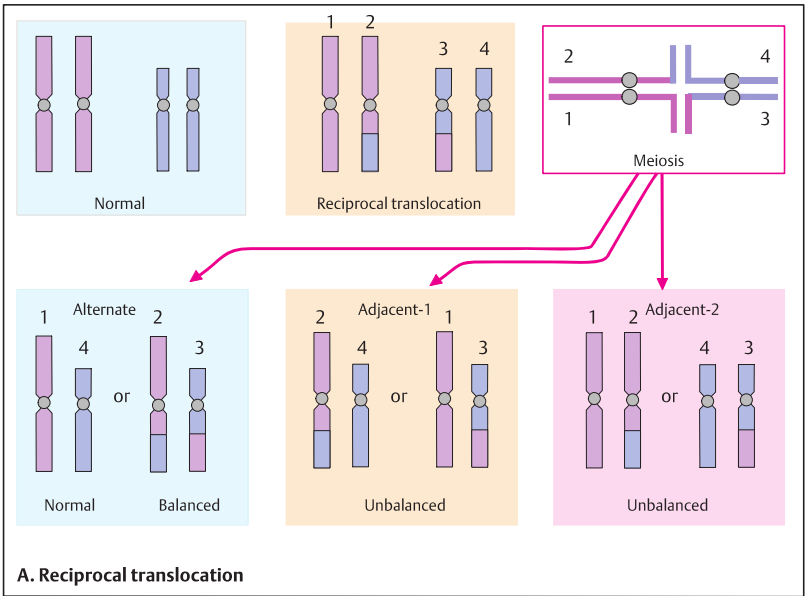
Gersen SL, Keagle MB, eds. The Principles of Clinical Cytogenetics. 2nd ed. Totowa, NJ: Humana Press, 2005

Harper P. Practical Genetic Counseling. 7th ed. London: Edward Arnold, 2010

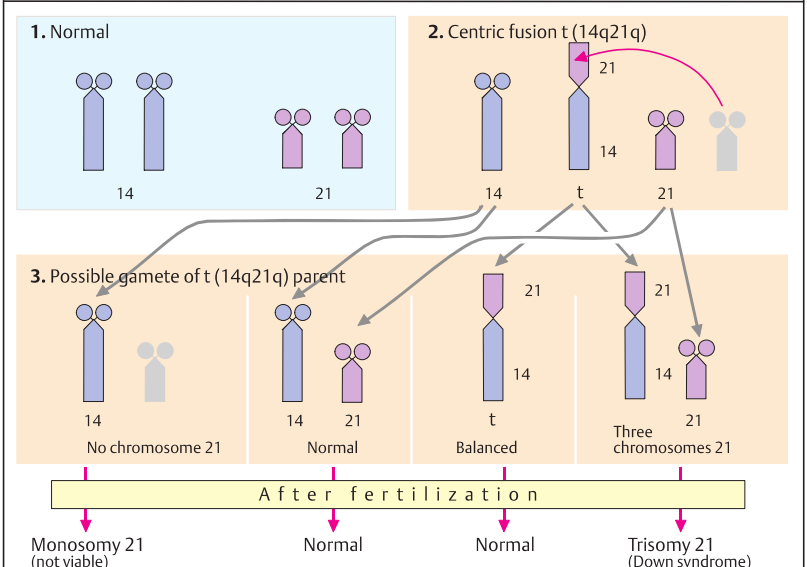
Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001

Rooney DE, Czepulski BH, eds. Human Cytogenetics. A Practical Approach. 2nd ed. Oxford: Oxford University Press, 2001

Schinzel A. Catalogue of Unbalanced Chromosome Aberrations in Man. 2nd ed. Berlin: De Gruyter, 2001



A. Reciprocal translocation



B. Centric fusion of acrocentric chromosomes

Structural Chromosomal Aberrations

Structural chromosome abnormalities result from one break or several breaks that disrupt the continuity of a chromosome. The break may occur at any stage of the cell cycle. If it occurs during the G_1 stage and remains unrepaired during S phase, it will involve both chromatids as a *chromosome break*, visible at the next metaphase stage. When the break occurs during G_2, it will be present in one chromatid of the next metaphase (*chromatid break*). The principal types of structural chromosomal aberration as observed in human individuals are *deletion*, *duplication*, *inversion*, *isochromosome*, and the special case of a *ring chromosome*.

A. Deletion, duplication, and isochromosome

Chromosomal deletion (deficiency) arises from a single break with loss of the distal fragment (terminal deletion, **1**) or from two breaks and loss of the intervening segment (interstitial deletion, **2**). In molecular terms, terminal deletions are not terminal because telomeric sequences do not contain genes (see p. 136). Duplications (**3**) occur mainly as small supernumerary chromosomes. Approximately half of these small chromosomes are derived from chromosome 15, being inverted duplications of the pericentric region (inv dup [15]). These represent one of the most common structural aberrations in man. An isochromosome (**4**) is an inverted duplication. It arises when a normal chromosome divides transversely instead of longitudinally and is then composed of two long arms or of two short arms. In each case, the other arm is missing.

B. Inversion

An inversion is a 180-degree change in direction of a chromosomal segment caused by a break at two different sites, followed by reunion of the inverted segment. Depending on whether the centromere is involved, a *pericentric* inversion (the centromere lies within the inverted segment) and a *paracentric* inversion can be differentiated.

C. Ring chromosome

A ring chromosome arises after two breaks with the loss of both ends, followed by joining of the two newly resulting ends. Since its distal segments have been lost, a ring chromosome is unbalanced.

D. Aneusomy by recombination

When a normal chromosome pairs with its inversion-carrying homolog during meiosis, a loop is created in the region of the inversion (**1**). When the inverted segment is relatively large, crossing-over may occur within this region (**2**). In the daughter cells, one resulting chromosome will contain a duplication of segments A and B and a deficiency of segment F (**3**), whereas the other resulting chromosome lacks segments A and B and has two segments F (**4**).

E. A ring chromosome at meiosis

A ring chromosome at mitosis and meiosis is often unstable and is frequently lost or duplicated. Ring chromosomes tend to generate new variants of derivative chromosomes, all imbalanced. If crossing-over occurs during meiosis, a dicentric ring can be created. At the following anaphase, the ring breaks at variable locations as the centromeres go to different poles. The daughter cells will receive different parts of the ring chromosome, resulting in deficiency in one cell and duplication in the other. Ring chromosomes may be dicentric.

Medical relevance

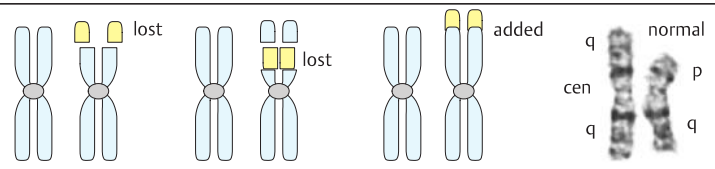
Structural chromosomal aberrations are an important group of human developmental disorders. Structural chromosomal rearrangements occur with a frequency of approximately 0.7 to 2.4 per 1,000 in mentally impaired individuals. Small supernumerary chromosomes are observed approximately once in 2,500 prenatal diagnoses.

Further Reading

Gu S, et al. Mechanisms for complex chromosomal insertions. PLoS Genet 2016; 12(11): e1006446

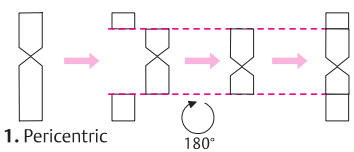
Madan K. Paracentric inversions: a review. Hum Genet 1995; 96(5): 503–515

Schinzel A. Catalogue of Unbalanced Chromosome Aberrations in Man. 2nd ed. Berlin: De Gruyter, 2001



1. Terminal deletion **2.** Interstitial deletion **3.** Duplication **4.** Isochromosome for the long arm (q) of the X chromosome

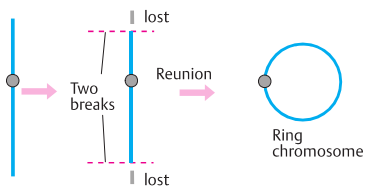
A. Deletion, duplication, isochromosome



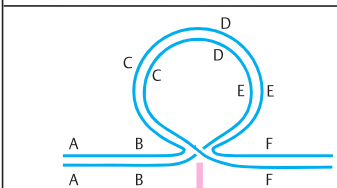
1. Pericentric

2. Paracentric

B. Inversion

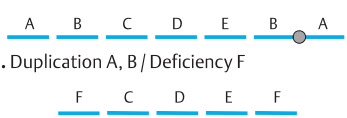


C. Ring chromosome



1. Inversion loop

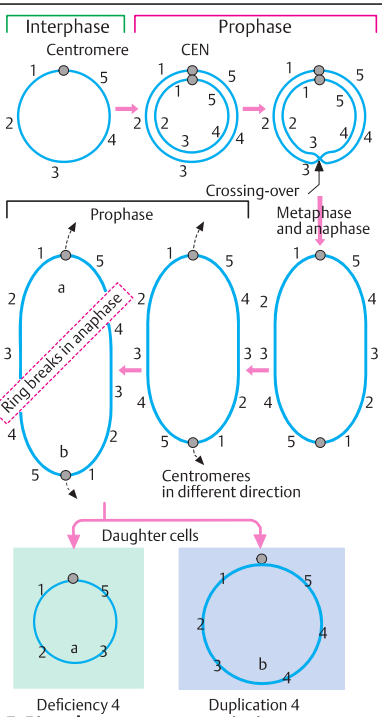
2. Crossing-over between C and D



3. Duplication A, B / Deficiency F

4. Deficiency A, B / Duplication F

D. Aneusomy by recombination



E. Ring chromosome at meiosis

Ribosomal RNA and Protein Assembly

Ribosomes are large ribonucleoprotein particles (RNPs) composed of two-thirds RNA and one-third proteins. Here the genetic information contained in messenger RNA (mRNA) and transfer RNA is decoded for protein synthesis. All proteins produced by a cell at any given time are called the *proteome*. The total number of different proteins required for eukaryotes is estimated to be 90,000. Ribosomes are the products of individual genes, the ribosomal genes.

A. Structure and components of ribosomes

Ribosomes in prokaryotes and eukaryotes consist of two subunits, a larger one and a smaller one. The size of a ribosome is expressed by its sedimentation coefficient (70 S in prokaryotes, 80 S in eukaryotes). The sedimentation coefficient is a measure of the rate of sedimentation in an ultracentrifuge of a molecule suspended in a less dense solvent; it is measured in Svedberg units (S); values are not additive.

The prokaryote small subunit is 30 S and the large subunit is 50 S in size. The prokaryote ribosome is composed of three different ribosomal RNA (rRNA) molecules (5, 35, and 165 S in size) and 34 proteins. A bacterial cell contains approximately 20,000 ribosomes, which account for approximately 25% of its mass. The 30-S subunit, containing a large 16-S rRNA and 21 proteins, is the site where genetic information is decoded. It also has a proofreading mechanism. The 50-S subunit provides peptidyl transferase activity. The entire ribosome has a molecular mass of 2.5 million daltons (MDa).

The larger eukaryotic ribosome (80 S, 4.2 MDa) consists of a 40-S subunit and a 60-S subunit. The 60-S subunit contains 5-, 5.8-, and 28-S rRNAs (120, 160, and 4,800 bases, respectively) in addition to 50 proteins. The 40-S subunit has 18-S rRNAs (1,900 bases) and 33 proteins. Bacterial 30- and 50-S ribosomal subunit structures have been recognized at 5-Å resolution.

B. From gene to protein

Transcription and processing of the primary transcript (splicing) takes place in the nucleus. RNA in the nucleus is bound to nuclear RNA-binding proteins for stabilization. The mature RNA is released from the nucleus into the cytoplasm, where it associates with ribosomes for translation into the corresponding sequence of amino acids.

C. Nucleolus and synthesis of ribosomes

The nucleolus is a morphologically and functionally specific region in the cell nucleus in which ribosomes are synthesized. In a male, the rRNA genes (200 copies per haploid genome) are transcribed by RNA polymerase I to form 45-S rRNA precursor molecules. After the 45-S rRNA precursors have been produced, they are rapidly packaged with ribosomal proteins (from the cytoplasm). Before they are transferred from the nucleus to the cytoplasm, they are cleaved to form three of the four rRNA subunits. These are released into the cytoplasm with the separately synthesized 5-S subunits. Here they form functional ribosomes. Two types of small RNAs have important functions. Small nuclear RNAs (snRNAs) are a family of RNA molecules that bind specifically with a small number of nuclear RNPs (snRNPs, pronounced “snurps”). These play important roles in the modification of RNA molecules after transcription (posttranscriptional modification). snRNAs base-pair with pre-mRNA and with each other during the splicing of RNA. Small nucleolar RNA molecules assist in processing pre-rRNAs and in assembly of ribosomes. (Figures adapted from Alberts al, 2015).

Medical relevance

Mutations in genes encoding small ribonucleoproteins cause different human disorders. A variety of chemical compounds occurring naturally as poisons or synthetic products are used for cancer therapy by inhibition of transcription or translation.

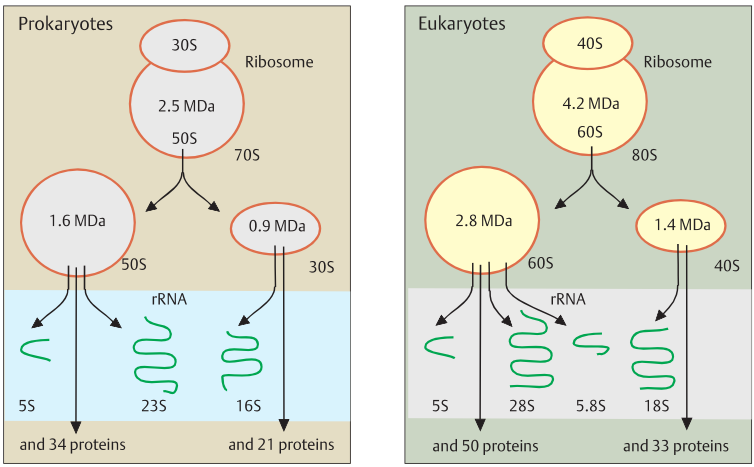
Further Reading

Agalarov SC, et al. Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain. Science 2000; 288(5463): 107–113

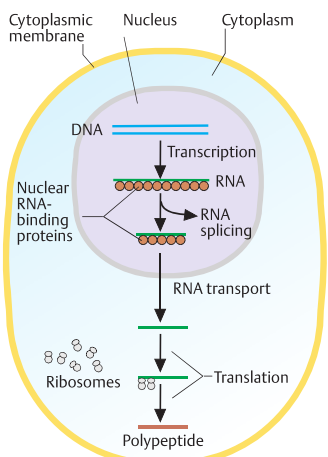
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Berg JM, et al. Biochemistry. 7th ed. New York, NY: W. H. Freeman, 2012

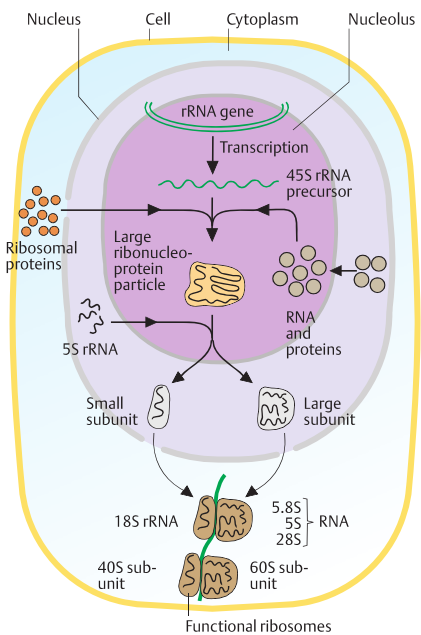
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016



A. Overview of the structure and components of ribosomes



B. From gene to protein



C. Nucleolus and synthesis of ribosomes

Stages of Transcription

Transcription (see p. 52) occurs in different stages. RNA polymerase II transcribes protein-coding genes and many noncoding RNAs. For this purpose, the DNA double helix is transiently opened, and the template strand is used to direct synthesis of an RNA strand. This process starts and ends at a defined site. The RNA strand is synthesized from the 5′ end to the 3′ end by RNA polymerase II.

A. Transcription by RNA polymerase II

Transcription occurs in four stages. The first step of transcription is *template recognition* with binding of RNA polymerase II to a specific sequence of DNA molecules known as the promoter (**1**). The double helix then opens, and an initiation complex makes the template strand available for base pairing. *Initiation* (**2**) begins with synthesis of the first RNA molecules at the initiation complex. Initiation requires several other proteins, which are collectively referred to as activators and transcription factors. *Elongation* (**3**) begins when the RNA polymerase enzyme moves along the DNA, thereby extending the RNA chain. As it moves, DNA helicases unwind the DNA double helix. The DNA that has been transcribed rewinds into the double helix behind the polymerase. At *termination* (**4**), the RNA polymerase is removed from the DNA. At this point, the formation of the unstable primary transcript is completed. Since it is unstable, it is immediately translated in prokaryotes and modified (processed) in eukaryotes (see p. 56). All processes are mediated by the complex interaction of a variety of enzymes (not shown).

B. Polymerase-binding site

The polymerase-binding site defines the starting point of transcription. At the termination site, it is closed again (rewinding). Bacterial RNA polymerase binds to a specific region of approximately 60 base pairs (bp) of the DNA.

C. Promoter of transcription

A promoter is a DNA sequence that specifies the site of RNA polymerase binding from which transcription is initiated. The promoter is organized into several regions with sequence homology. In eukaryotes, transcription of protein-coding genes begins at multiple sites, often extending hundreds of base pairs upstream. A specific DNA sequence of approximately 4 to 8 bp is located approximately 25 to 35 bp upstream (in the 5′ direction) of the gene. Since this sequence is nearly the same in all organisms, it is called a *consensus sequence*. One such sequence is called the *TATA box* because it contains the sequence TATA (or T)AA(or T)TA/G. It is highly conserved in evolution. In prokaryotes, the promoter contains a consensus sequence of 6 bp, TATAAT (also called a *Pribnow box* after its discoverer), located 10 bp above the starting point. Another region of conserved sequences, TTGACA, is located 35 bp upstream of the gene. These sequences are referred to as the **–**10 box and the **–**35 box, respectively.

D. Transcription unit

A transcription unit is the segment of DNA between the sites of initiation and termination of transcription.

E. Determining transcription start sites

A segment of DNA suspected to contain a transcription start site can be analyzed using the nuclease S1 protection assay. Endonuclease S1, an enzyme present in the mold *Aspergillus oryzae*, cleaves single-stranded DNA and RNA, but not double-stranded DNA or RNA. Therefore, a DNA fragment that is denatured and mixed with total RNA from cells containing the gene to be analyzed will hybridize with cognate RNA and form a DNA/RNA heteroduplex. Following digestion with S1, all single-stranded DNA will be removed, allowing the relevant DNA to be identified.

Medical relevance

Changes in regulatory DNA sequences are causes of human genetic disorders.

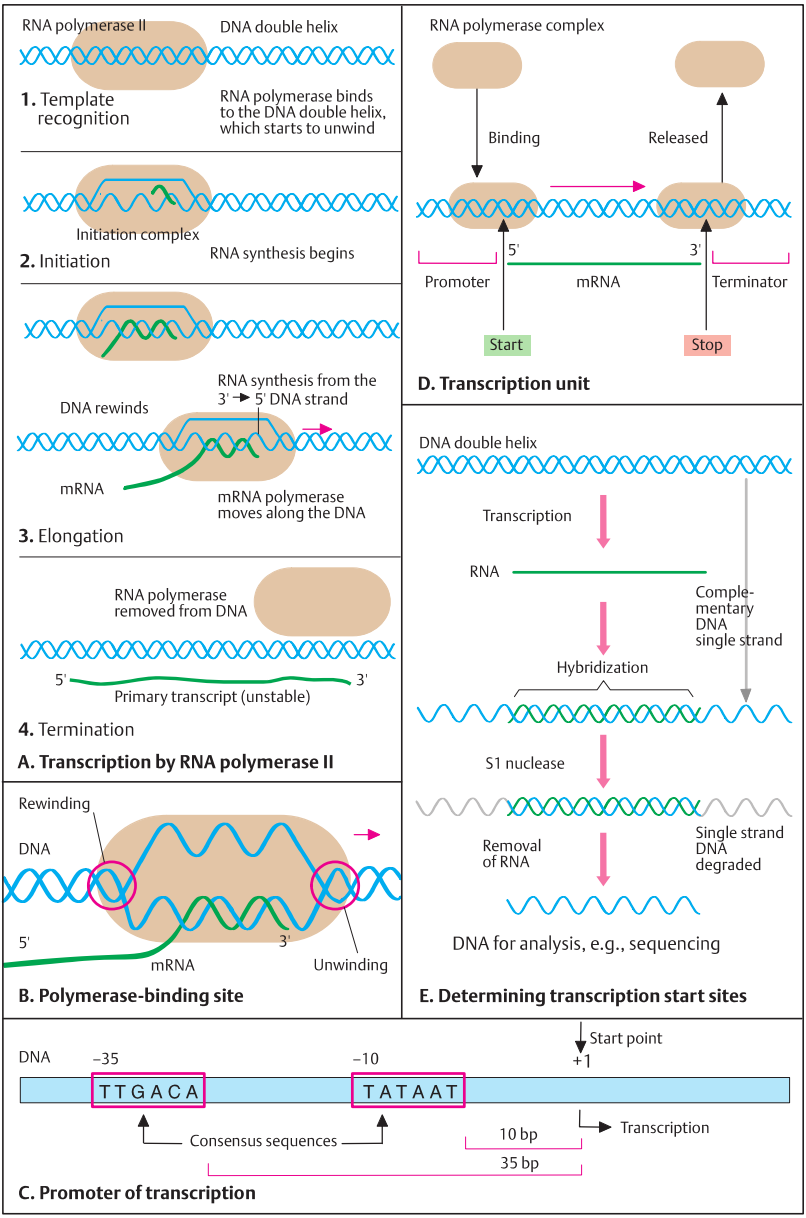
Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015


RNA polymerase II
DNA double helix
1. Template recognition
RNA polymerase binds to the DNA double helix, which starts to unwind
Initiation complex
2. Initiation
RNA synthesis begins
DNA rewinds
RNA synthesis from the 3' → 5' DNA strand
mRNA
mRNA polymerase moves along the DNA
3. Elongation
RNA polymerase removed from DNA
5'
3'
Primary transcript (unstable)
4. Termination
A. Transcription by RNA polymerase II
Rewinding
DNA
5'
3'
mRNA
Unwinding
B. Polymerase-binding site
RNA polymerase complex
Binding
Released
5'
3'
Promoter
mRNA
Terminator
Start
Stop
D. Transcription unit
DNA double helix
Transcription
RNA
Comple-mentary DNA single strand
Hybridization
S1 nuclease
Removal of RNA
Single strand DNA degraded
DNA for analysis, e.g., sequencing
E. Determining transcription start sites
DNA
–35
–10
Start point
+1
TTGACA
TATAAT
Consensus sequences
10 bp
35 bp
Transcription
C. Promoter of transcription

Basic Principles of Gene Control

Regulatory DNA sequences are regions that bind to DNA-specific proteins (transcription factors [TFs]) and regulate gene activity. Two types are distinguished: promoters and enhancers. The binding of TFs at the promoter is the most important mechanism for regulating gene activity.

A. Consensus sequences at the promoter

The DNA sequence at promoters is well conserved in evolution. Owing to their functional importance, they do not tolerate mutational changes. In prokaryotes, two important regulatory sequences are 35 and 10 nucleotide base pairs upstream (in the 5′ direction) of the starting point of transcription (Pribnow, 1975). The **–**10 site (TATA box) is 5′-TATAAT-3′. The **–**35 sequence is 5′-TATT-GACA-3′ (**1**). Mutations at different sites upstream of a gene have different effects, depending on their contribution to transcription at a precise location (**2**). (Figure adapted from Rosenberg & Court,1979).

B. Assembly of general transcription factors

In eukaryotes, the TATA box in the promoter region is located approximately 25 to 35 base pairs (bp) upstream of the transcription start site (**1**). Several other promoter-proximal elements (different TFs) are involved in regulating gene activity. General TFs associate in an ordered sequence. First, TFIID (transcription factor D for polymerase II) binds to the TATA region (**2**). The TATA box is recognized by a small, 30-kDa TATA-binding protein (TBP), which is part of one of the many subunits of TFIID (the bending of the DNA by TBP is not shown here). Then TFIIB binds to the complex (**3**). Subsequently, other TFs (TFIIH, followed by TFIIE) and RNA polymerase II (Pol II), escorted by TFIIF, join the complex and ensure that Pol II is attached to the promoter (**4**). Pol II is activated by phosphorylation, and transcription can begin (**5**). Other activities of TFIIH involve a helicase and an ATPase (adenosine triphosphatase). The site of phosphorylation is a polypeptide tail, in which the serine (S) and threonine (T) side chains are phosphorylated.

C. RNA polymerase promoters

Eukaryotic cells contain three RNA polymerases (Pol I, Pol II, and Pol III). Each uses a different type of promoter. RNA polymerase II is a 550-kDa complex of 12 subunits. It requires a TF complex (TFIID, see **B**) that binds to a single upstream promoter (**1**). RNA polymerase I has a bipartite promoter, one part 170 to 180 bp upstream (5′ direction) and the other from approximately 45 bp upstream to 20 bp downstream (3′ direction). The latter is called the *core promoter* (**2**). Pol I requires two ancillary factors: UPE1 (upstream promoter element 1) and SL1. RNA polymerase III uses either upstream promoters or two internal promoters downstream of the transcription start site (**3**). Three TFs are required with internal promoters: TFIIIA (a zinc finger protein, see p. 170), TFIIIB (a TBP and two other proteins), and TFIIIC.

Pol I is located in the nucleolus and synthesizes ribosomal RNA. It accounts for approximately 50 to 70% of the relative activity. Pol II and Pol III are located in the nucleoplasm (the part of the nucleus excluding the nucleolus). Pol II represents 20 to 40% of cellular activity. It is responsible for the synthesis of heterogeneous nuclear RNA, the precursor of mRNA. Pol III is responsible for the synthesis of transfer RNAs and other small RNAs. It contributes only minor activity of approximately 10%. Each of the large eukaryotic RNA polymerases (500 kDa or more) has 8 to 14 subunits and is more complex than the single prokaryotic RNA polymerase.

Further Reading

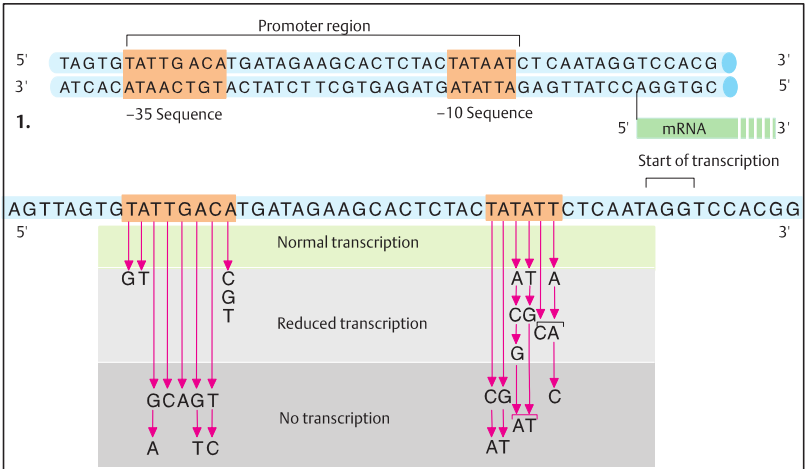
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

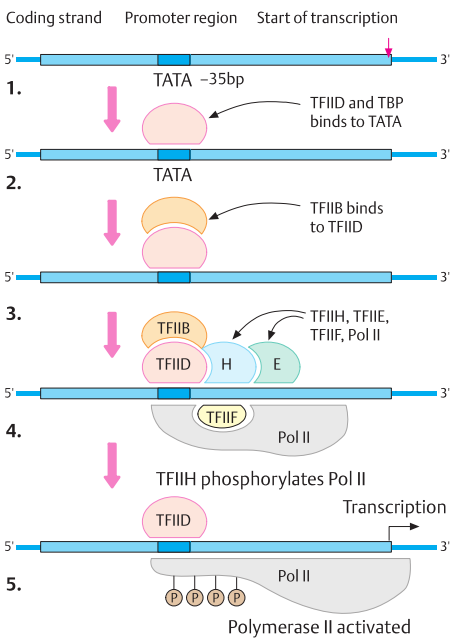
Pribnow D. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proc Natl Acad Sci U S A 1975; 72(3): 784–788

Rosenberg M, Court B. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu Rev Genet 1979; 13: 319–353

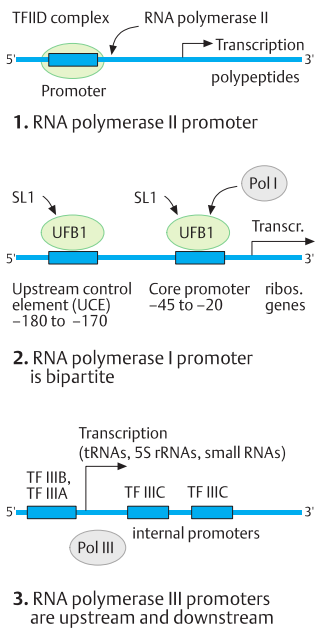


2. Effect of mutations in the promoter region on the rate of transcription

A. Consensus sequences at the promoter



B. Assembly of general transcription factors



C. RNA polymerase promoters

Regulation of Gene Expression in Eukaryotes

The term *gene expression* refers to the entire process of decoding the genetic information of active genes. When genes are active (expressed) throughout the life of a cell or an organism, they show *constitutive expression*. Those genes that are transcribed only under certain circumstances, in specific cells or at specific times, show *conditional expression*.

A. Levels of control

Schematically, one can distinguish four levels at which gene expression in eukaryotes is regulated at. The first, and by far the most important, is primary control of the initiation of transcription. Second, processing of the transcript to mature messenger RNA (mRNA) can be regulated at the level of the primary RNA transcript by alternative splicing (see **D**). Third, gene control is possible at the level of translation by mRNA editing (see **B**). Finally, at the protein level, posttranslational modifications can determine the activity of a protein. A newly recognized form of gene expression control is RNA interference (see p. 174).

B. RNA editing

RNA editing modifies genetic information at the RNA level. An important example is the gene encoding apolipoprotein (Apo) B-100, which is involved in lipid metabolism (OMIM 107730). It encodes a 512-kDa protein of 4,536 amino acids. This is synthesized in the liver and secreted into the blood, where it transports lipids. Apo B-48 (250 kDa), a functionally related shorter form of the protein with 2,152 amino acids, is synthesized in the intestine. An intestinal deaminase converts a cytosine in codon 2,152 CAA (glutamine) to uracil (UAA). This change results in a stop codon (UAA) and thereby terminates translation at this site.

C. Long-range gene activation by an enhancer

The term *enhancer* refers to a DNA sequence that stimulates the initiation of transcription. Enhancers act at a distance from the gene. They may be located upstream or downstream on the same DNA strand (*cis*-acting) or on a different DNA strand (*trans*-acting). Enhancer effects are mediated by sequence-specific DNA-binding proteins. One model suggests that DNA forms a loop between an enhancer and the promoter. An activator protein bound to the enhancer, for example, a steroid hormone, could then come into contact with the general transcription factor complex at the promoter. Enhancer elements provide tissue-specific or time-dependent regulation.

D. Alternative RNA splicing

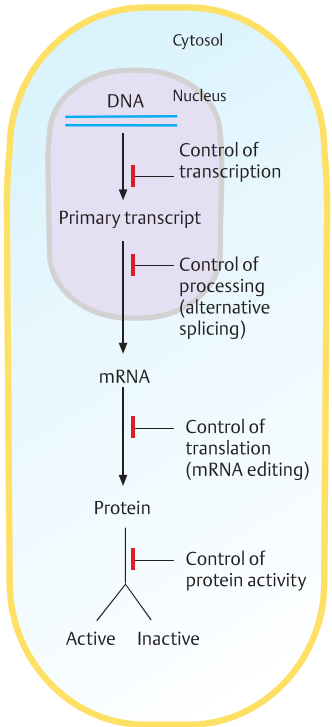
Different forms of mRNA are usually obtained from the same gene by alternative splicing. This is an important mechanism for generating multiple protein isoforms from a single gene. The resulting proteins differ slightly in their amino acid sequence. This may result in small functional differences. Quite often, these differences are restricted to certain tissues, as schematically shown here for the calcitonin gene (OMIM 114130). The primary transcript for the calcitonin gene contains six exons. They are spliced into two different types of mature mRNA. One type, calcitonin, consisting of exons 1 to 4 (excluding exons 5 and 6), is produced in the thyroid. The other, consisting of exons 1, 2, 3, 5, and 6 and excluding exon 4, encodes a calcitonin-like protein in the hypothalamus (calcitonin gene related product).
Alternative splicing clearly represents an evolutionary advantage because it allows for a high degree of functional flexibility.

Medical relevance

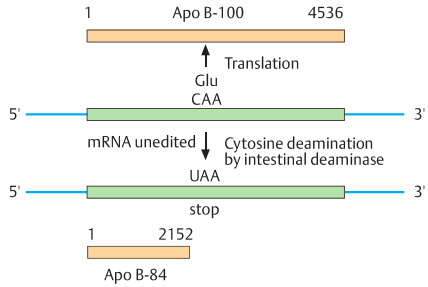
Human genetic disorders may result from defective splicing or DNA changes in regulatory sequences (promoters and enhancers).

Further Reading

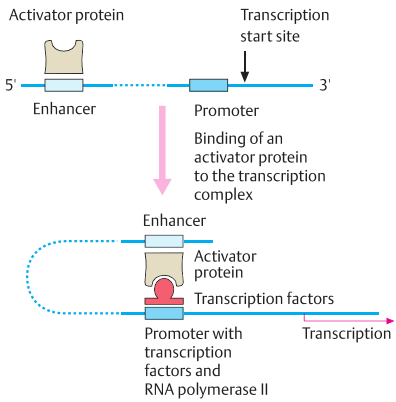
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015
Graveley BR. Alternative splicing: increasing diversity in the proteomic world. Trends Genet 2001; 17(2): 100–107
Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016
Modrek B, Lee CA. A genomic view of alternative splicing. Nat Genet 2002; 30(1): 13–19
OMIM. Mendelian Inheritance of Man. Available at www.ncbi.nlm.nih.gov/omim



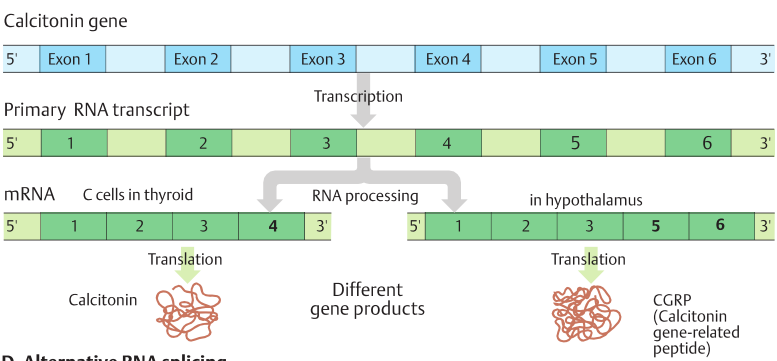
A. Levels of control of eukaryotic gene expression



B. RNA editing



C. Long-range gene activation by an enhancer



D. Alternative RNA splicing

DNA–Protein Interactions

Regulatory DNA sequences exert their control function by interaction with specific DNA-binding proteins. Regulatory proteins serving as transcription factors recognize specific DNA sequences by a precise fit of their surface to the DNA double helix.

A. Binding of a regulatory protein to DNA

Gene regulatory proteins recognize DNA sequence information without affecting the hydrogen bonds within the helix. Each base pair represents a distinctive pattern of hydrogen bond donors (shown in the rectangle in red) and hydrogen acceptors (shown in the rectangle in green). These proteins bind to the major groove of DNA. A single contact of an asparagine (Asn) residue of a gene-regulatory protein with a DNA base adenine (A) is shown here. A typical area of surface-to-surface contact involves 10 to 20 such interactions, resulting in high specificity.

B. Specific DNA–protein interaction

The α-helix of a DNA-binding regulatory protein recognizes specific DNA sequences. Many bacterial repressor proteins are dimeric, and therefore an α-helix from each dimer can insert itself into two adjacent major grooves of the DNA double helix (recognition or sequence-reading helix). This structural motif is called a *helix–turn–helix* motif because two helices lie next to each other. The example shows the tight interaction of the bacteriophage 434 repressor protein with one side of the DNA molecule over a length of 1.5 turns.

C. Zinc finger transcription factors

Several eukaryotic regulatory proteins harbor regions that fold around a central zinc atom (Zn^{2+}) to form a structural motif resembling a finger (hence, zinc finger). The basic zinc finger motif consists of a zinc atom connected to four amino acids of a polypeptide chain. The three-dimensional structure on the right consists of an antiparallel β-sheet (amino acids 1–10), an α-helix (amino acids 12–24), and the zinc connection. Four amino acids—two cysteines at positions 3 and 6 and two histidines at positions 19 and 23—are bonded to the zinc atom and hold the carboxy (COOH) end of the α-helix to one end of the β-sheet.

D. Zinc finger binding to DNA

The α-helix of each zinc finger can contact the major groove of the DNA double helix and establish specific and strong interactions over several turns in length. Extraordinary flexibility in gene control has been acquired during evolution through adjustments to the number of interacting zinc fingers. Zinc finger proteins serve important functions during embryonic development and differentiation.

E. Hormone response element

Some DNA-binding proteins act as signal transmitting molecules. The signal may be a hormone or a growth factor that activates an intracellular receptor. Steroid hormones enter target cells and bind to specific receptor proteins. Three views of the glucocorticoid receptor and its binding to DNA are shown here. The dimeric glucocorticoid receptor consists of two polypeptide chains (*hormone response element*). Each is stabilized by a zinc ion connected to four cysteine side chains (**1**). The skeletal model shows the binding of the dimeric protein to the DNA double helix (**2**). The space-filling model (**3**) shows how tightly the helix of each dimer of this protein (brown above, red below) fits into two neighboring major grooves of DNA (shown in red and green). (Figures adapted from Alberts et al, 2015, and Lodish et al, 2016).

Medical relevance

Defective function of transcription factors causes certain developmental disorders, for example, Waardenburg syndrome type 1 (OMIM 193500), Waardenburg syndrome type 2A (OMIM 193510) by mutations in MITF (microphthalmia-associated transcription factor; OMIM 156845), or cancer (see *TP53* gene, p. 318).

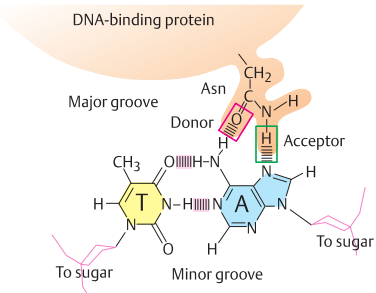
Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

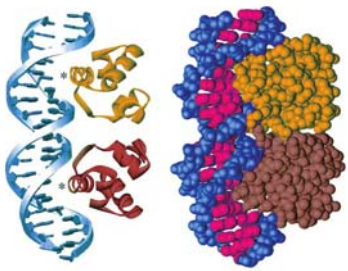
Berg JM, et al. Biochemistry. 7th ed. New York, NY: W. H. Freeman, 2012

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

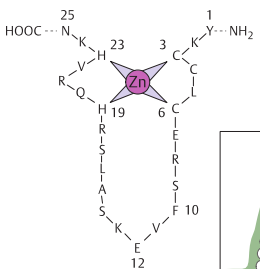
OMIM. www.ncbi.nlm.nih.gov/omim

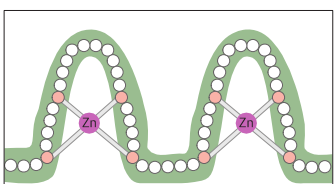


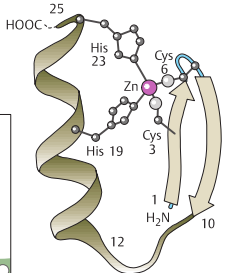
A. Binding of a regulatory protein to DNA



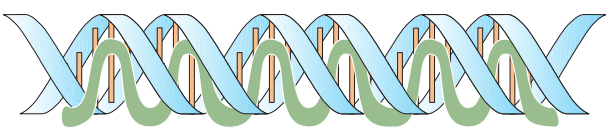
B. Interaction of a DNA-binding protein with DNA



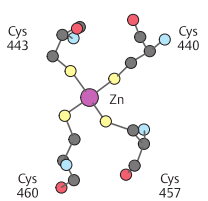




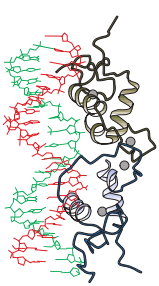
C. Zinc finger motif



D. Zinc finger binding to DNA



1.



2.

3.

E. Binding to a response element

Other Forms of Transcription Control

A number of DNA-binding proteins control eukaryotic transcription as activators or repressors. They are characterized by different structural motifs. The specific DNA–protein binding usually involves noncovalent interactions between atoms in an α-helix of the binding domain of the protein and atoms on the outside of the major groove of the DNA double helix. Approximately 2,000 transcription factors are encoded by the human genome. Examples of important classes of DNA-binding domains are (i) *homeodomain proteins* containing a 180-base-pair (bp) sequence that is highly conserved as a *homeobox* in evolution (see section on Genes in Embryonic Development, p. 198), (ii) *zinc finger proteins* (see the previous page), (iii) *leucine-zipper proteins* (see below), and (iv) *basic helix–loop–helix* (bHLH) proteins (see the previous page).

A. Leucine zipper and basic helix–loop–helix protein

Most DNA-binding regulatory proteins have a dimeric structure, which allows a dual function: one part of the molecule recognizes specific DNA sequences, and the other stabilizes the molecule. One frequently occurring class of proteins has a characteristic structural motif called a *leucine zipper*. The name is derived from its basic structure. A typical leucine zipper protein consists of a dimer with a periodic repeat of leucine every seven residues. The leucine residues are aligned along one face of each α-helix and interact with the DNA at adjacent major grooves (**1**). In bHLH proteins, the DNA-binding helices at the N-terminal near the DNA are separated by nonhelical loops (**2**).

B. Alternative combinations

Leucine zipper proteins and bHLH proteins often exist in alternative combinations of dimers consisting of two different monomers. This dramatically increases the number of combinatorial possibilities. In higher eukaryotes, leucine zipper proteins often mediate the effect of cyclic adenosine monophosphate (cAMP) on transcription. Genes under this type of control contain a cAMP response element (CRE), which is a palindromic 8-bp recognition sequence. A protein of 43 kDa binds to this target sequence. It is therefore known as the CRE-binding protein. Leucine zipper proteins were first described in 1988 by Landschulz et al., in 1988.

C. Activation by steroid hormone

Transcriptional enhancers are regulatory regions of DNA that increase the rate of transcription. Binding of a hormone receptor complex to a specific DNA sequence, a hormone response element, activates an enhancer. This activates the promoter, and transcription begins. Numerous genes in mammalian development are regulated by steroid-responsive transcription.

D. Detection of DNA–protein interactions

DNA–protein interactions can be detected by a DNase I footprinting assay (**1**). This is based on the observation that DNA is protected from digestion by nucleases at the protein-binding site, whereas DNase I outside the binding site digests DNA. The protected protein-binding sites of DNA appear as missing bands (footprint) after separation of the DNA fragments by gel electrophoresis. The electrophoretic mobility shift assay, or band-shift assay (**2**), is based on the principle that a DNA–protein complex retards the speed at which a fragment migrates in gel electrophoresis. (Figures based on Alberts et al, 2015, and Lodish et al, 2016).

Medical relevance

Mutations in a bHLH protein-coding gene, *TWIST* (OMIM 601622), cause acrocephaly–syndactyly syndrome type III (Saethre–Chotzen syndrome, OMIM 101400).

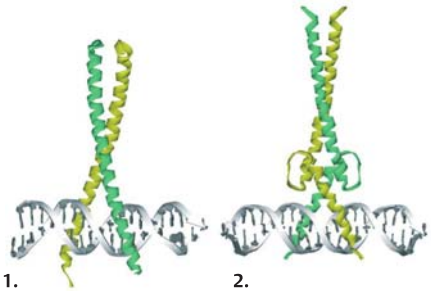
Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

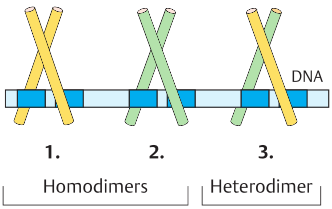
Landschulz WH, et al. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 1988; 240(4860): 1759–1764

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

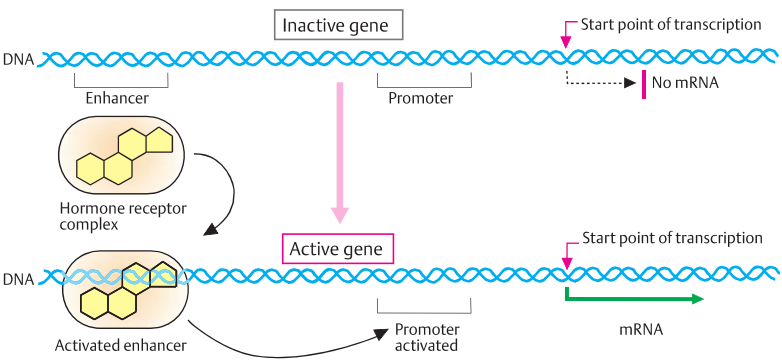
OMIM. Mendelian Inheritance of Man. Available at www.ncbi.nlm.nih.gov/omim



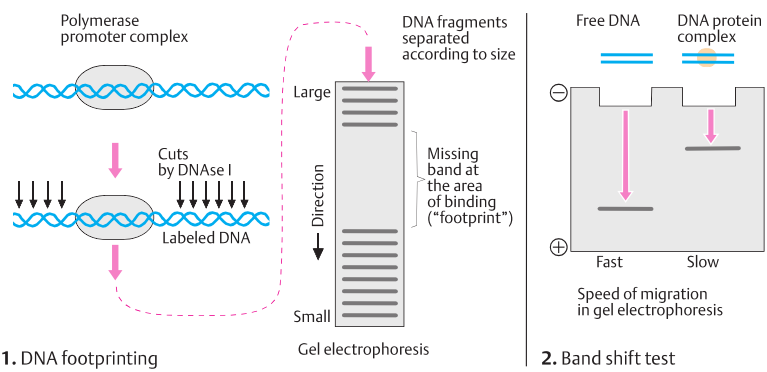
A. Leucine zipper and bHLH protein



B. Alternative heterodimeric combinations



C. Activation by steroid hormone receptor complex binding to an enhancer



1. DNA footprinting

2. Band shift test

D. Detection of DNA–protein interaction

Noncoding RNAs

Small noncoding RNAs regulate gene expression and protect the genome from invading foreign RNA molecules. Different types can be distinguished on the basis of their structure and function: microRNAs (miRNAs) and short interfering RNAs (siRNAs). A third type, Piwi-associated RNAs are restricted to germ cells. More than 1,000 different miRNAs regulate approximately one-third of human protein-coding genes. siRNAs serve in a defense function and gene control by inducing mRNA degradation or translational repression (RNA interference, RNAi).

A. Short interfering RNA

RNAi, discovered in 1978, is induced by siRNA. This typically consists of 21 to 23 nucleotides of double-stranded RNA (dsRNA) with a 2-bp overhang at both ends, and with high specificity for the nucleotide sequence of the target molecule: an mRNA.

B. RNA-induced silencing complex

siRNA interacts with proteins that have helicase and RNA nuclease activity, resulting in an RNA-induced silencing complex (RISC). RISC is a sequence-specific nuclease with helicase and RNase activity to cleave RNA. The helicase unwinds the dsRNA, and the RNA endonuclease cleaves it.

C. Degradation of double-stranded RNA

As part of defense system, foreign dsRNA (**1**), for example, from some RNA viruses, is bound to a double-stranded specific protein complex with helicase and RNA endonuclease activity, called a *dicer* (**2**). dsRNA is unwound by the helicase activity. The neighboring single-stranded RNA (red arrows) is degraded by a specialized ribonuclease III in the RISC (**3**). The resulting short fragments (red arrows) are then rapidly degraded by cellular nucleases (**4**).

D. Transcriptional gene silencing

Transcriptional gene silencing targets mRNA by RNAi. A protein complex forms an RNA-induced transcriptional silencing complex. It binds complementary RNA transcripts as they emerge from RNA polymerase II and cleaves it, resulting in continued repression of the target gene.

E. Functional effect of RNAi

RNAi can be used for intentional silencing of a selected gene to assess its normal function. In this example, a gene in the developing worm *Caenorhabditis elegans* (see p. 202) is silenced. dsRNAi targeted to a specific gene was produced by inserting the transcript into a plasmid vector in two orientations, sense and antisense (**1**). The noninjected embryo developed normally (**2a**). The injected dsRNA on the other hand interfered with normal mRNA production (**2b**). The degrading effect of the dsRNA is visualized by fluorescence in situ hybridization of a labeled probe of mRNA from the target gene. The probe hybridizes to the normal, noninjected embryo (purple on the left, **2a**) but not to the injected embryo (**2b**), whose target gene mRNA has been destroyed. (Figures in A-D modified from McManus & Sharp, 2002, and Kitabwalla & Ruprcht, 2002; in E from Lodish et al, 2016).

Medical relevance

Since RNAi can selectively silence any gene, possibly one with a harmful mutation, it holds promise for future developments toward use in therapy.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Fire A, et al. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 1998; 391(6669): 806–811

Ghildiyal M, et al. Small silencing RNAs: an expanding universe. Nat Rev Genet 2009; 10(2): 94–108

Grosshans H, et al. Molecular biology: the expanding world of small RNAs. Nature 2008; 451(7177): 414–416

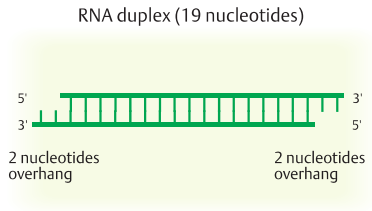
Kitabwalla M, Ruprecht RM. RNA interference–a new weapon against HIV and beyond. N Engl J Med 2002; 347(17): 1364–1367

Lim LP, et al. Vertebrate microRNA genes. Science 2003; 299(5612): 1540

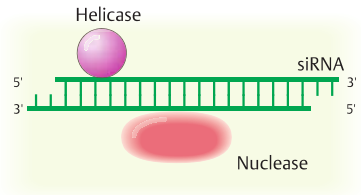
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

McManus MT, Sharp PA. Gene silencing in mammals by small interfering RNAs. Nat Rev Genet 2002; 3 (10): 737–747

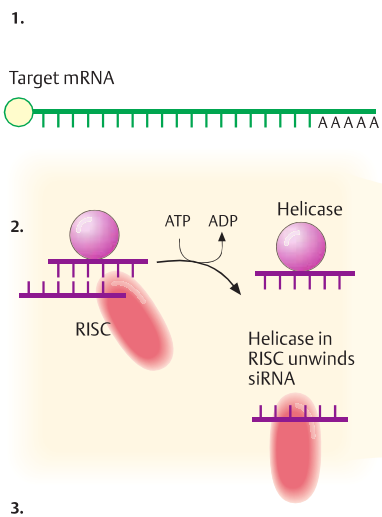
Strachen T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015



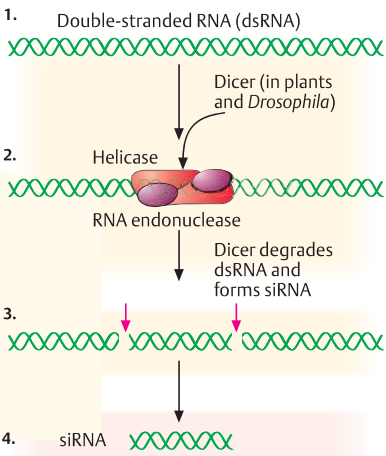
A. Short interfering RNA (siRNA)



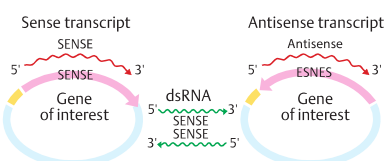
B. RNA-induced silencing complex (RISC)



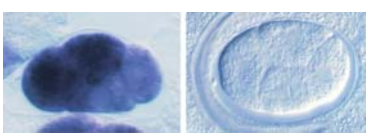
D. Transcriptional gene silencing



C. Degradation of double-stranded RNA (dsRNA) by dicer



1. In vitro production of double-stranded RNA



2a. Noninjected **2b.** Injected

E. Functional effect of RNAi

Targeted Gene Disruption

Targeted gene disruption refers to experimental inactivation of a gene to investigate its function in higher organisms. In a "knockout" animal, usually mice, the gene under study is inactivated in the germline by disrupting it (gene knockout). The effects can be studied at different embryonic stages and after birth. This knowledge can be used to understand the effects of mutations in homologous human genes as seen in human genetic diseases. A variant of knockout is known as gene *knock-in*. In this case, the targeting construct contains a normal gene that is introduced either in addition to or instead of the gene to be studied.

A. Preparation of embryonic stem cells with a knockout mutation

The target gene is disrupted (knocked-out) in yeast or embryonic stem (ES) cells of a mouse by homologous recombination with an artificially produced nonfunctional allele. The isolation of ES cells with a disrupted gene requires positive and negative selection.

To the DNA from the target gene (**1**), a bacterial gene conferring resistance to neomycin (neo^R) is added (**2**). In addition, DNA containing the thymidine kinase gene (tk^+) from herpes simplex virus is added to the gene replacement construct outside the region of homology and introduced into ES cells (**3**). Since the selective medium contains the positive and the negative selectable markers, neomycin and ganciclovir (an analogue of the nucleotide guanine), cells following nonhomologous recombination (**4**) and homologous recombination (**5**) containing the targeted gene can be distinguished. Random insertion in other genes (**6**) results in cells which die (**8**). Gene-targeted insertion (**7**), although rare, allows cells to grow in the selective medium and multiply (**9**).

Nonrecombinant cells and cells with nonhomologous recombination at random sites cannot grow in this medium because they remain sensitive to neomycin, whereas recombinant cells are resistant (positive selection). The gene encoding thymidine kinase (tk^+) confers sensitivity to ganciclovir. Since nonhomologously recombinant ES cells contain the tk^+ gene at random sites, they are sensitive to ganciclovir and cannot grow in its presence (negative selection). Only cells that have undergone homologous recombination can survive because they contain the gene for neomycin resistance (neo^R) and do not contain the tk^+ gene (**5**).

B. Transgenic mouse

ES cells from a mouse blastocyst are isolated (**1**) after 3.5 days of gestation (of a total of 19.5 days) and transferred to a cell culture grown on a feeder layer of irradiated cells that are unable to divide (**2**). ES cells heterozygous for the knockout mutation are added (**3**). They will be intergraded at a homologous site, although rarely. These ES cells are derived from a mouse that is homozygous for a different coat color (e.g., black) from that of the mouse that will develop from the blastocyst (e.g., white). The recombinant ES cells are injected into the recipient blastocyst (**4**). The early embryos are transplanted into a pseudopregnant mouse (**5**). Those offspring that have taken up ES-derived cells are chimeric, that is, consisting of normal cells and cells with an altered gene. The transgenic mice can be recognized by black coat color spots on a white (or brown) background (**6**). The chimeric mice are then backcrossed to homozygous white mice (**7**). Black offspring from this mating are heterozygous for the disrupted (mutant) gene (**8**). By further breeding of the heterozygous mice (**9**), some of their offspring, the knockout mice, will be homozygous for the disrupted gene, as visualized by their new coat color. (Figures adapted from Alberts et al, 2015, and Lodish et al, 2016).

Medical relevance

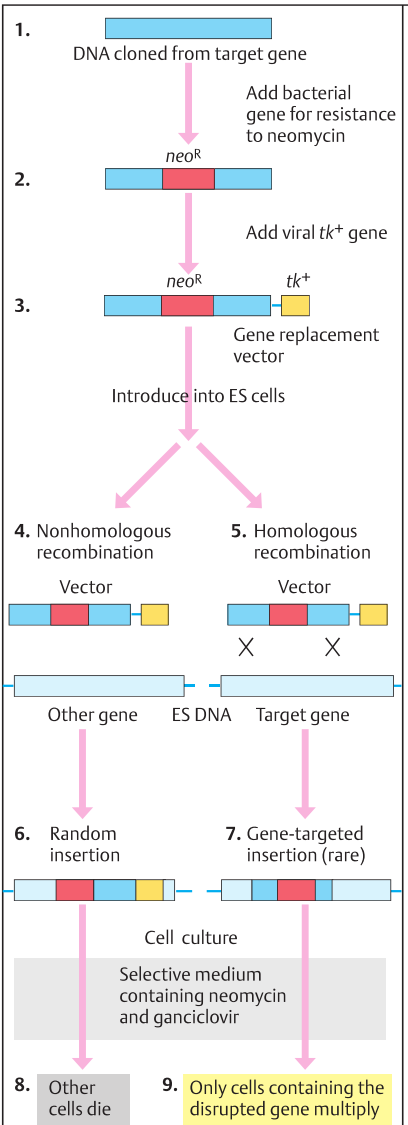
Comparison of the genotype and phenotype of a knockout mouse with a corresponding human genetic disease may yield information about the effects of a mutation, in particular during embryonic development.

Further Reading

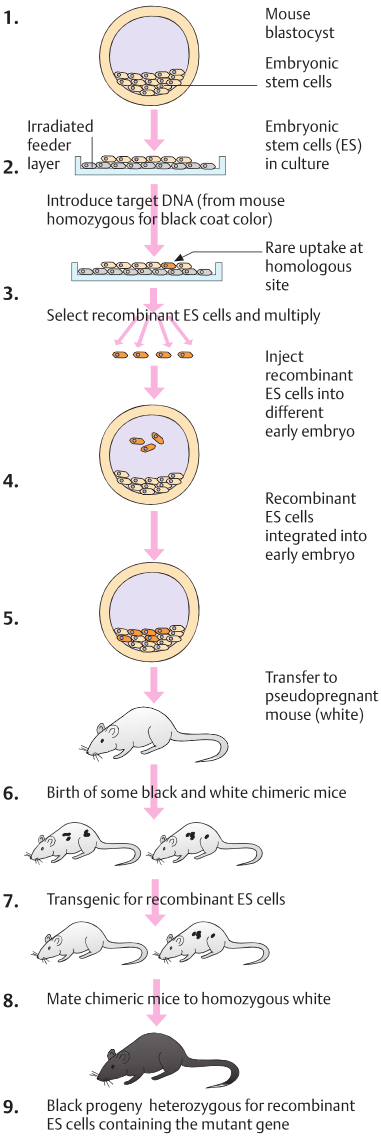
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Capecchi MR. Targeted gene replacement. Sci Am 1994; 270(3): 52–59

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016



A. Preparation of ES cells with a knockout mutation



B. Transgenic mice with targeted gene disruption

DNA Methylation

Methylation of DNA influences the functional state of regulatory regions without changing the nucleotide sequence, referred to as *epigenetic* modification. Methylated regions are not transcribed, whereas regions free of methylation are present in promoters and enhancers of expressed genes. The site of methylation is cytosine. In position 5, the H atom can be replaced by a methyl group to yield 5-methylcytosine (5-meC; see p. 76). Heritable methylation occurs at CpG dinucleotides. Certain enzyme pathways set, maintain, or erase DNA methylation.

A. Maintenance of DNA methylation

DNA methyltransferase 1 (DNMT1, OMIM) maintains the methylation pattern in somatic cells. It adds methyl groups to the newly synthesized DNA strand. The methylated sites in the parental DNA (**1**) serve as templates for correct methylation of the two new strands after replication (**2**). This ensures that the previous methylation pattern is correctly maintained at the same sites as the parental DNA (**3**).

B. *De novo* DNA methylation

DNMT3A (OMIM 602769) and DNMTA3B (OMIM 602900) add methyl groups at new positions to both strands of DNA. Unmethylated DNA (**1**) is methylated by these enzymes in a site-specific and tissue-specific manner (**2**). Targeted homozygous disruption of the mouse *DNMT3A* and *DNMT3B* genes results in severe developmental defects. DNMT3C protects male germ cells from transposon activity (not shown).

C. Recognition of methylated DNA

Certain restriction enzymes are methylation-sensitive (**1**). The enzyme *Hpa*II cleaves DNA only when its recognition sequence 5′-CCGG-3′ is not methylated (**2**). In contrast, *Msp*I recognizes the same 5′-CCGG-3′ sequence irrespective of methylation (**3**). This difference can be used to distinguish the methylation pattern for diagnostic purposes. Other methods identify methylated regions by applying sodium bisulfite, which deaminates cytosine to uracil, whereas 5-meC remains unchanged.

D. Human *DNMT3B* gene and mutations

Mutations in the human gene *DNMT3B* encoding type 3B de novo methyltransferase causes ICF syndrome type 1 (immunodeficiency, centromeric chromosomal instability, and facial anomalies; OMIM 242860). The centromeres of chromosomes 1, 9, and 16 are unstable. The human *DNMT3B* gene (**1**) consists of 23 exons spanning 47-kb DNA. Six exons are subject to alternative splicing. The protein (**2**) has 845 amino acids with five DNMT motifs (I, IV, V, IX, X) in the C-terminal region. The arrows point to six different mutations. One mutation (**3**) changes an A into a G in codon 809, that is, GAC (Asp) to GGC (Gly), which results in a replacement of asparagine (Asn) by glycine (Gly). Both parents are heterozygous for this mutation. Multiradiate chromosomes with multiple p and q arms are typical in lymphocytes in ICF syndrome, here derived from chromosomes 1 and 16 shown in R banding (**4**). (Figure adapted from Xu et al, 2005).

Medical relevance

Aside from the example given in part D, aberrant methylation patterns resulting from mutations encoding enzymes responsible for methylation and demethylation are involved in several developmental abnormalities and tumors.

Further Reading

Barau J, et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 2016; 354(6314): 909–912

Hagleitner MM, et al. Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). J Med Genet 2008; 45 (2): 93–99

Hansen RS, et al. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A 1999; 96(25): 14412–14417

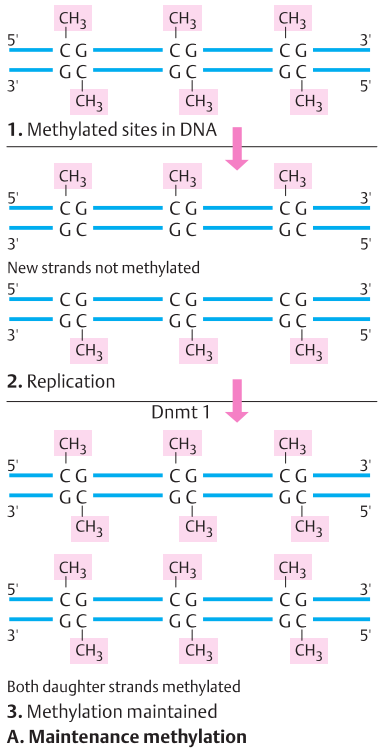
Robertson KD. DNA methylation and human disease. Nat Rev Genet 2005; 6(8): 597–610

Schübeler D. Function and information content of DNA methylation. Nature 2015; 517(7534): 321–326

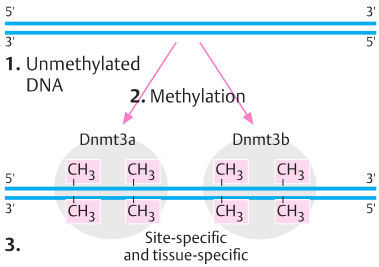
Smith ZD, et al. DNA methylation: roles in mammalian development. Nat Rev Genet 2013; 14(3): 204–220

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015

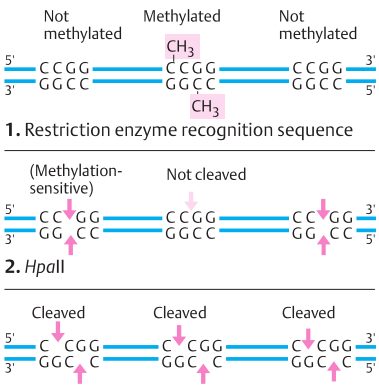
Xu GL, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 1999; 402(6758): 187–191



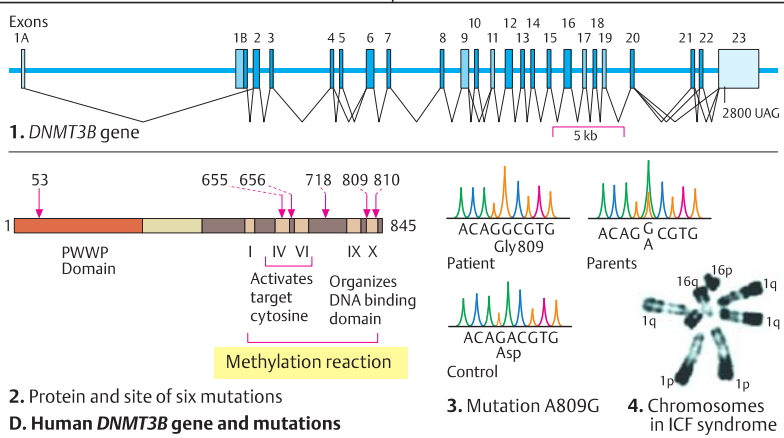
A. Maintenance methylation



B. De novo DNA methylation



C. Recognition of a methylated DNA segment



D. Human *DNMT3B* gene and mutations

Reversible Changes in Chromatin Structure

Chromatin can assume one of two alternative stages: open (*euchromatin*), associated with an active transcriptional state, or closed (*heterochromatin*), inactive. In differentiated cells, the local transcriptional state is permanent, whereas in other regions, it is reversible. Reversible changes (chromatin remodeling) are key events in gene regulation. The local structure of chromatin is an *epigenetic* state. Chromatin remodeling is an active, reversible process by which histones of the nucleosomes are displaced from the DNA molecule to make genes accessible for replication or transcription.

A. Histone modifications

Multiple changes in chromatin structure occur in the N-terminal tails of histones H3 and H4. There are several sites for epigenetic modification by methylation, acetylation (adding acetyl groups), or phosphorylation. Examples are shown for histones H3 (**1**) and H4 (**2**). Trimethylation (Me3) of lysine (K) in H3 in position 4 (H3K4Me3) and acetylation (Ac) in position 27 (H3K27Ac) is associated with active promoters, H3K4Me1 (one CH_3 group) and H3K27Ac occurs with active enhancers.

Active chromatin is acetylated at the lysine residues of H3 and H4 histones. Inactive chromatin is methylated in H3 at the lysine position 9 (H3K9Me3) and at other lysine residues (H3K27Me3) Lys-9 in at other lysine residues. Thus, multiple modifications can occur and influence each other. Combinations of different signals are called the *histone code*. (Data based on Strachan et al., 2015, and B. Horsthemke, pers. communication.)

B. Histone acetylation and deacetylation

Acetylation of core histones occurs during DNA replication and activation of certain genes. Major targets are lysines in the N-terminal tails of H3 and H4. Active genes are acetylated directly in nucleosomal histones. Acetylation is mediated by histone acetyltransferases, which are part of a large activating complex (**1**). The acetyl groups can be removed in a reversible process, deacetylation (**2**), mediated by deacetylases (histone deacetylases), which are associated with repressors. Acetylases are associated with gene activators. Deacetylation and methylation may be connected. Two methylcytosine-binding proteins, MeCP1 and MeCP2, selectively bind to methylated DNA. Transcriptional repression by the methyl-CpG-binding proteins 1 and 2 involves histone deacetylation in a multiprotein complex. (Figures in B and C adapted from Lodish et al, 2016).

C. Chromatin remodeling

Several different activator proteins can reverse the "gene-off" state of condensed heterochromatin. These are specific DNA-binding control elements that are able to interact with multiprotein complexes. The activator proteins bind to a mediator protein. As a result, chromatin becomes decondensed and the gene assumes the "gene-on" state. General transcription factors and RNA polymerase assemble at the promoter and initiate transcription.

Medical relevance

Dysregulated chromatin states form a newly recognized group of chromatin disorders, (see p. 238). Mutations of the *MECP2* gene on Xq28 (OMIM 300005) cause Rett syndrome (OMIM 312750).

Further Reading

Amir RE, et al. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 1999; 23(2): 185–188

Bickmore WA, et al. Perturbations of chromatin structure in human genetic disease: recent advances. Hum Mol Genet 2003; 12(Spec No 2): R207–R213

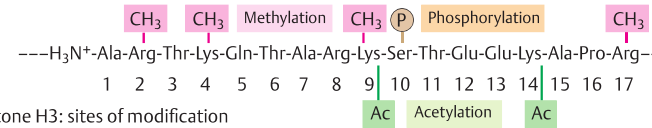
Chahrour M, et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. Science 2008; 320(5880): 1224–1229

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

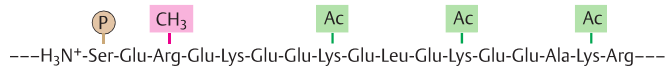
Moog U, et al. Neurodevelopmental disorders in males related to the gene causing Rett syndrome in females (MECP2). Eur J Paediatr Neurol 2003; 7 (1): 5–12

Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature 2010; 465(7299): 721–727

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015



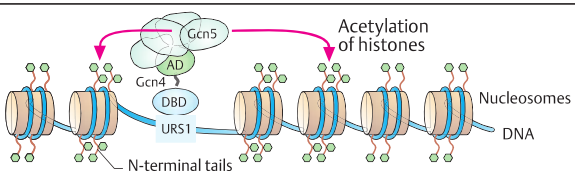
1. Histone H3: sites of modification



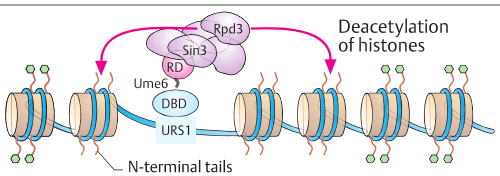
2. Histone H4: sites of modification

A. Histone modification – a key event in gene regulation

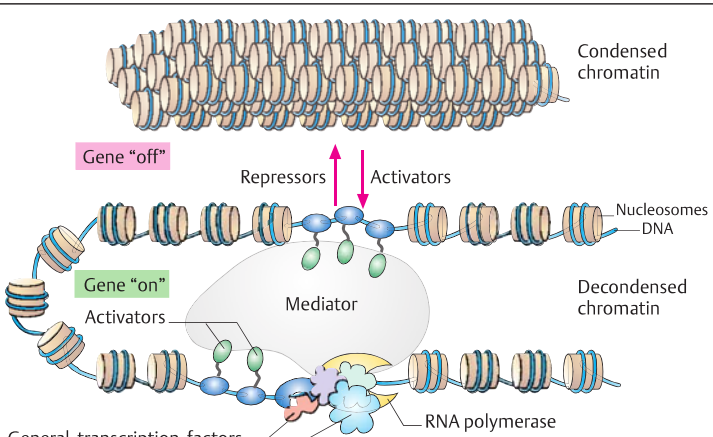
1. Activator-directed histone acetylation



2. Repressor-directed histone deacetylation



B. Histone acetylation and deacetylation



C. Chromatin remodeling in transcription control

Genomic Imprinting

Genomic imprinting is an epigenetic process leading to parent-of-origin specific gene expression of just one allele, but not the other. It occurs in placental mammalian and marsupial genomes only. The inherited imprint pattern is maintained during cell division. It is erased and reestablished in the primordial germ cells according to the sex of the developing embryo. Imprinting is assumed to have evolved in mammals in response to intrauterine competition for resources (parental conflict hypothesis). Natural selection acts differently on maternal and paternal genomes. Presumably, it favors a balance between preserving maternal resources and fetal growth. Approximately 75 imprinted genes in the human genome are known. Most occur in clusters where they are regulated by imprinting control regions.

A. The importance of two different parental genomes

In the 1980s, it was shown that normal development in mice requires a contribution from both parental genomes. When from a normal zygote (**1**) the pronucleus of female origin is replaced by another male pronucleus (androgenote, **2**), it fails to complete preimplantation. The rare few that reach postimplantation develop abnormally and do not progress beyond the 12-somite stage. In contrast, a normal zygote develops normally (**3**). If the male pronucleus is replaced by another female pronucleus (gynogenote, **4**), approximately 85% of gynogenotes develop normally until preimplantation. However, subsequently, the extraembryonic membranes do not develop properly and the embryo dies at or before the 40-somite stage. (Figure adapted from Sapienza & Hall, 2001).

B. Requirement for a maternal and a paternal genome

A naturally occurring human androgenetic zygote is a hydatidiform mole (**1**). This is an abnormal placenta containing two sets of paternal chromosomes and none from the mother. Implantation takes place, but an embryo does not develop. The placental tissues develop many cysts (**2**). In a naturally occurring gynogenetic zygote, when only maternal chromosomes are present, an ovarian teratoma with many different types of fetal tissue develops (**3**). No placental tissue is formed. In triploidy, a relatively frequent fetal human chromosomal disorder (see p. 382), extreme hypoplasia of the placenta and fetus is observed when the additional chromosomal set is of maternal origin (**4**, 69,XXX). (Photographs provided by Professor Helga Rehder, Marburg and Wien).

C. Normal and Disturbed Imprint Patterns

The epigenetic changes responsible for imprinting occur in early embryogenesis. The imprint pattern typically presents in somatic cells according to parental origin of the chromosomes. The normal female imprint pattern is shown by a red square (**1**). It is erased in primordial germ cells. It is reestablished in germ cells. Thus, the zygote receives the proper imprint pattern. This is maintained through all subsequent cell divisions. Failure of imprint erasure (**2**) results in an erroneous imprint pattern. This results in developmental disorders, Prader–Willi syndrome, and Angelman syndrome (see Imprinting Disorders, p. 258). Failure of imprint establishment (**3**) or imprint maintenance (**4**) likewise results in an aberrant imprint pattern. Such errors may result in familial occurrence of these two developmental disorders. (Figure adaptd from Horsthemke, 2010).

Medical relevance

Failure in imprint erasure, establishment, or maintenance causes an important heterogeneous group of *imprinting disorders* (see p. 258).

Further Reading

Barton SC, et al. Role of paternal and maternal genomes in mouse development. Nature 1984; 311 (5984): 374–376

Catalogue of Parent of Origin Effects. University of Otago. Available at http://igc.otago.ac.nz/home.html

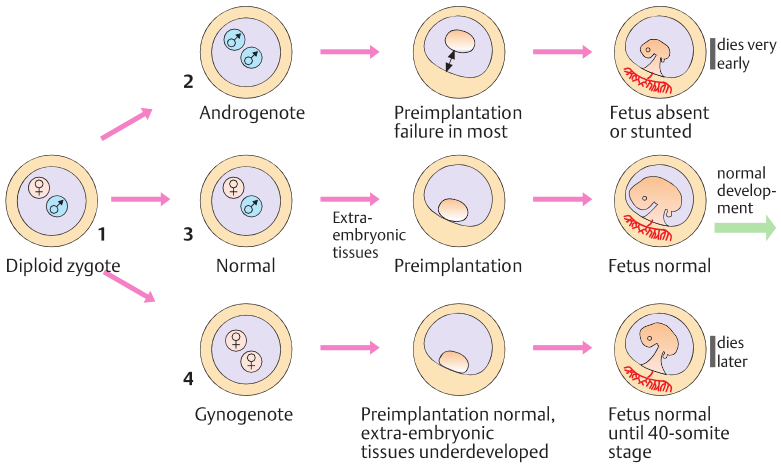
Ferguson-Smith AC. Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet 2011; 12(8): 565–575

Horsthemke B. Mechanisms of imprint dysregulation. Am J Med Genet C Semin Med Genet 2010; 154C(3): 321–328

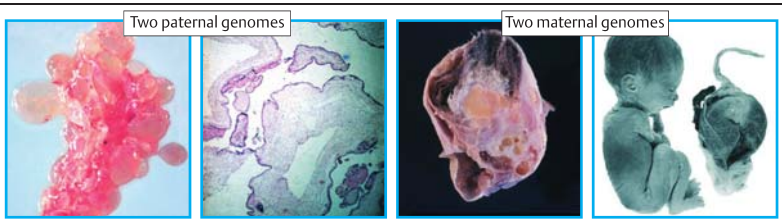
Peters J. The role of genomic imprinting in biology and disease: an expanding view. Nat Rev Genet 2014; 15(8): 517–530

Reik W, et al. Genomic imprinting: parental influence on the genome. Nat Rev Genet 2001; 2(1): 21–32

Sapienza C, Hall JG. Genetic imprinting in human disease, pp. 417-431. In: Scriver CR et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York: McBraw-Hill, 2001

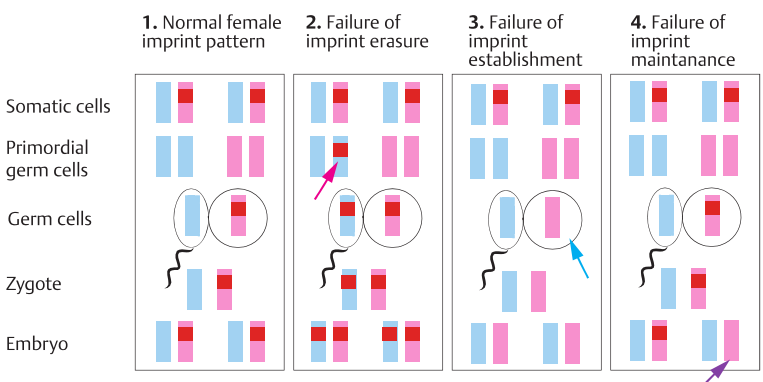


A. The importance of two different parental genomes



1. Hydatidiform mole **2.** Hydatidiform mole **3.** Ovarian teratoma **4.** Triploidy 69,XXX

B. Human embryonic development depends on presence of a maternal and a paternal genome



C. Normal and disturbed imprint patterns

Mammalian X Chromosome Inactivation

In mammals, most genes on one of the two X chromosomes in every cell of females are inactivated at random. This results in *dosage compensation*, which equalizes the dosage of X-chromosomal gene products in XX and XY cells. X-inactivation is initiated from an X-inactivation center, a *cis*-acting region on the X chromosome under the control of the *XIST* gene (OMIM 314670) located at Xq13.2. In early embryonic development, a noncoding RNA (Xist, X-inactivation specific transcript) silences one of the X chromosomes in cells with two or more X chromosomes by methylation and histone modifications.

A. X chromatin

Small, darkly staining bodies were described by Barr and Bertram in 1949 in the nerve cells of female cats (**1, 3**), but not in males (**2**). In 1954, Davidson and Smith described similar structures as drumsticks in peripheral human blood leukocytes (**4**). X chromatin is visible as a dark density in the nuclei of oral mucosal cells (**5**).

B. Scheme of X inactivation

The maternal *Xist* gene is expressed from the morula stage onward. It follows a random permanent X-inactivation pattern involving either the maternal or the paternal X chromosome, stably transmitted to all daughter cells.

C. Mosaic pattern of expression

In 1961, Mary F. Lyon described a mosaic distribution pattern of X-linked coat colors in female mice as a manifestation of X inactivation (**1**). Fingerprints of human females heterozygous for X-linked hypohidrotic ectodermal dysplasia (OMIM 305100) also show a mosaic pattern of areas with normal sweat pores (black points) and areas without normal sweat pores as in affected males (**2**). In cell cultures from females heterozygous for X-chromosomal HGPRT (hypoxanthine-guanine phosphoribosyltransferase) deficiency (OMIM 308000), the colonies are either $HGPRT^-$ or $HGPRT^+$ (**3**).

D. X-inactivation profile

Genes in defined regions of the human X chromosome are not activated. An X-inactivation profile established by Carrel and Willard in 2005 revealed that 458 (75%) genes are inactivated and 94 (15%) regularly escape inactivation. In addition, 65 genes (10%) are inactivated in some females, but not in others. Thus, 25% of X-linked human genes are not regularly inactivated, and 10% exhibit an interindividual inactivation pattern. To the left (**a**) of the X chromosome, nine vertical lanes represent nine rodent/ human cell hybrids. Genes expressed on the inactive X chromosome are shown in blue; silenced genes are shown in yellow. The right (**b**) illustrates the level of expression in the inactive X chromosome.

E. Evolutionary strata on the X

The human X chromosome harbors strata (S1–S5) of different evolutionary origin and time (see p. 230).

Medical relevance

Deviation from the random choice of either the maternal or paternal X chromosome to be inactivated may lead to mild clinical manifestations in heterozygous females for an X-chromosomal mutation (OMIM 300087).

Further Reading

Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 2005; 434(7031): 400–404

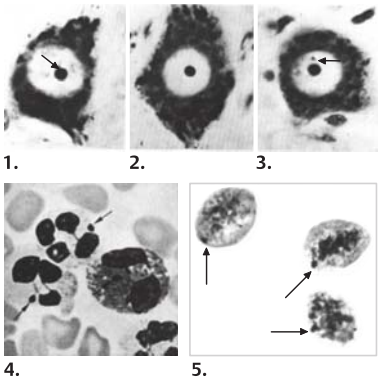
Cerase A, et al. Xist localization and function: new insights from multiple levels. Genome Biol 2015; 16: 166

Lyon MF. Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 1961; 190: 372–373

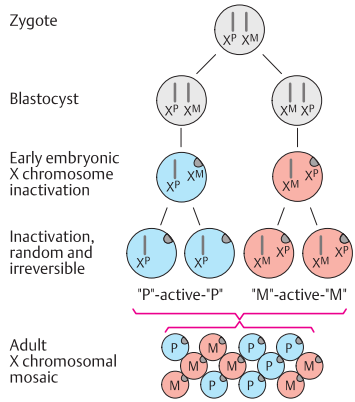
Migeon BR. Studies of skin fibroblasts from 10 families with HGPRT deficiency, with reference in X-chromosomal inactivation. Am J Hum Genet 1971; 23(2): 199–210

Passarge E, Fries E. X chromosome inactivation in X-linked hypohidrotic ectodermal dysplasia. Nat New Biol 1973; 245(141): 58–59

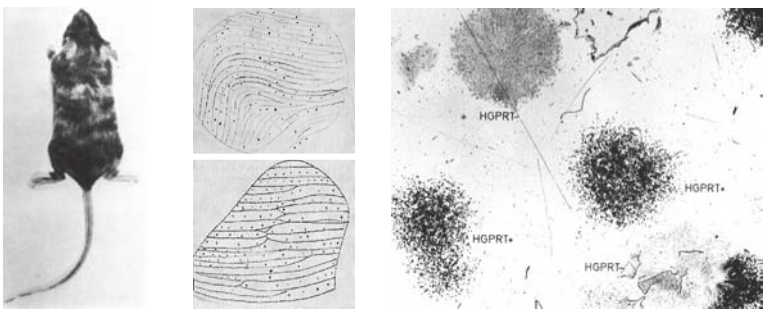
Thompson MW. Genetic implications of heteropyknosis of the X chromosome. Can J Genet Cytol 1965; 7: 202–213



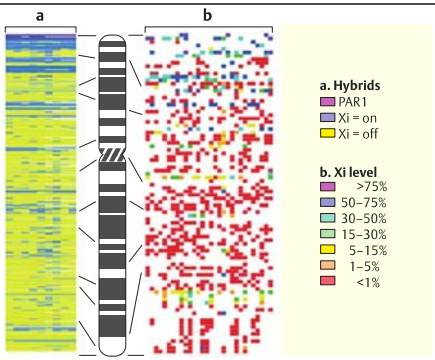
A. X chromatin



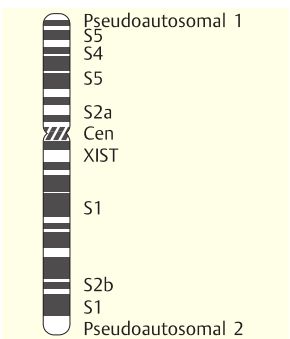
B. Scheme of X inactivation



C. Mosaic pattern of expression of X-linked genes



D. X-inactivation profile of the human X chromosome



E. Strata on the X chromosome

Cellular Signal Transduction

Multicellular organisms use a broad repertoire of signaling molecules for communication between and within cells (see p. 86). The specific binding of an extracellular signaling molecule (ligand) to its receptor on a target cell triggers a specific response. This is followed by a series of mutually activating or inhibitory molecular events, called a *signal transduction pathway* (or signaling pathway).

A. Main intracellular signal functions

Growth factors are important signaling molecules comprising a large group of secreted proteins (**1**). Each binds with high specificity to a cell surface receptor protein (**2**). This activates intracellular signal transduction proteins (**3**) and initiates a cascade of activations of responsive proteins (often by phosphorylation) that act as second messengers (**4**). Hormones are small signaling molecules (**5**) that arrive through the bloodstream. They enter the cell either by diffusion or by binding to a cell surface receptor (**6**). Some hormones bind to an intranuclear receptor (**7**). Activated transcription factors (**8**) together with cofactors initiate transcription (**9**). Prior to transcription, an elaborate system of DNA damage recognition and repair mechanisms (**10**) checks DNA integrity (cell cycle control, **11**). Cell division proceeds if faults in DNA structure have been repaired; if not, the cell is sacrificed by apoptosis (cell death, **12**). (Figure adapted from Lodish et al, 2016).

B. Receptor tyrosine kinase family

Receptor tyrosine kinases (RTKs) are a major class of cell surface receptors. They consist of a single transmembrane protein with an extracellular N-terminal part, a transmembrane part, and an intracellular C-terminal part. The intracellular domain contains the tyrosine kinase domain. RTK ligands are growth factors that control a wide variety of functions in growth and differentiation. RTK receptors share structural features but differ in function. The extracellular, ligand-binding domains of RTKs contain cysteine-rich regions. In certain RTKs, the binding domains resemble immunoglobulin (Ig) chains (Ig-like domains), which are known for their ability to bind other molecules (see p. 296). The nine members of the RTK family shown from left to right are as follows:

EGFR (epidermal growth factor, OMIM 131550) and its ligands transmit signals in a variety of cellular functions, such as cell proliferation, differentiation, motility, and survival.

IR (insulin receptor, OMIM 147670, see p. 276).

FGFR 1, 2, 3 (fibroblast growth factor receptor 1, OMIM 136350; FGFR2, OMIM 176943; FGFR3, OMIM 134934; see p. 340), a family of receptors for fibroblast growth factors (FGF) involved in cell proliferation, differentiation, skeletal development, angiogenesis, and other functions.

PDGFR (platelet-derived growth factor receptors, type A, OMIM 173490; type B, OMIM 173410; and type C, OMIM 608452) are receptors for platelet-derived factors (PDGF) required for angiogenesis and blood vessel functions.

RET (rearranged during transformation, OMIM 164761, see p. 250).

v-erbB (a viral oncogene, a homolog to the human *ERBB2* gene, OMIM 164870). (Figure adapted from Brinvanlou & Darnell, 2002).

Medical relevance

Mutations in genes encoding RTKs may result in a proliferative signal in the absence of a growth factor and cause errors in embryonic development and differentiation (congenital malformations) or cancer. RTK mutations cause a group of important human diseases and malformation syndromes. The phenotypes due to the mutations differ according to the particular type of RTK involved and the type of mutation.

Further Reading

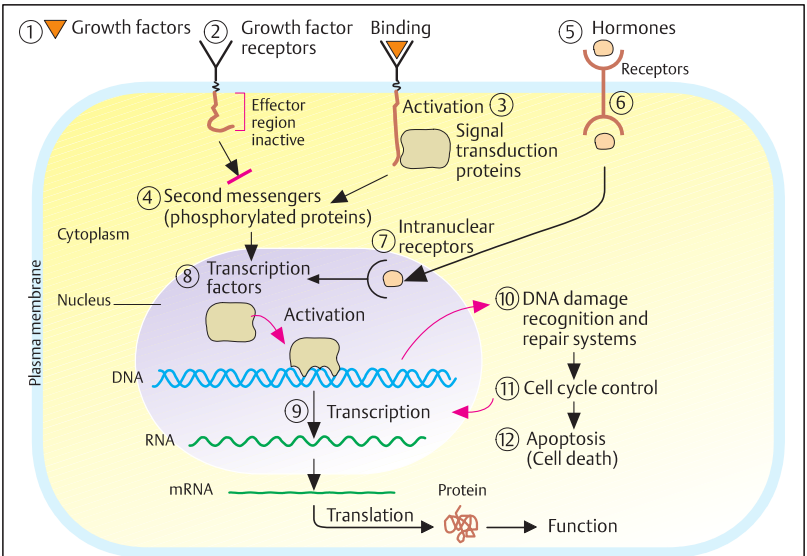
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Brivanlou AH, Darnell JEJr. Signal transduction and the control of gene expression. Science 2002; 295 (5556): 813–818

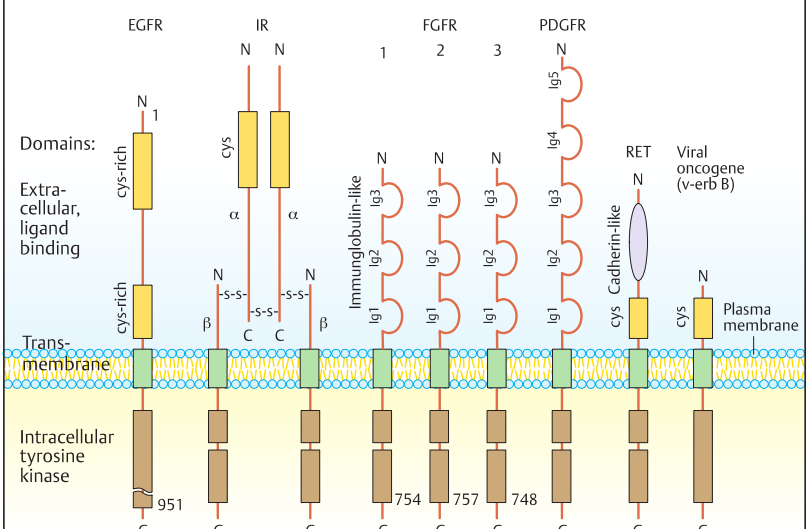
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Robertson SC, et al. RTK mutations and human. Trends Genet 2000; 16: 265-271

Tata JR. One hundred years of hormones. EMBO Rep 2005; 6(6): 490–496



A. Main intracellular signal functions



B. Receptor tyrosine kinase family

Heterotrimeric G Proteins

Heterotrimeric guanosine triphosphate (GTP) binding proteins, or just G proteins, are a family of guanine nucleotide binding proteins involved in transmitting extracellular signals into the cell by means of a G protein coupled receptor (GPCR). Binding of its ligand to the receptor on the surface of a responding cell causes a conformational change in the structure of the receptor. This induces a cascade of intracellular activations and inhibitions of proteins, which transmit the signal to the nucleus. There it induces or inhibits the transcription of a target gene. Activation may be achieved by inhibiting an inhibitor. Signal transduction by GPCRs mediates most external signals such for vision, taste, and smell (see p. 370).

A. Receptor tyrosine kinase

The binding of a ligand activates a receptor tyrosine kinase (see the previous page) by dimerization of two neighboring transmembrane proteins. As a result, intracellular tyrosine residues are phosphorylated. This activates another target protein and further signal transduction events.

B. G protein coupled receptors

G proteins are trimeric guanine nucleotide binding proteins with subunits, α, β, and γ. They act as molecular switches between an inactive and a brief active state. In their inactive state, the α subunit is bound to guanosine diphosphate (GDP). Upon binding of its ligand, it is activated, and the α subunit binds to GTP. This will induce a physiological effect by binding to an effector molecule. The α subunit is a GTPase. It rapidly hydrolyzes the bound GTP to GDP. As a result, the α subunit and the β γ complex reunite and form the inactive G protein. Therefore, the dissociated, active state of the α subunit and the β γ complex is short-lived. If the rapid reversal to the inactive state is delayed or impossible because of a toxin or a mutation, normal function is severely impaired. Approximately half of all known drugs cause their effects through GPCRs.
The α subunits of mammalian G proteins form a large family of signaling molecules that bind to a wide variety of effector proteins. Approximately 25 mammalian α subunits, 5 β subunits, and 12 γ subunits have been identified. GPCRs traverse the cell membrane seven times back and forth (therefore they are also called *serpentine receptors*).

C. Cyclic adenosine monophosphate

G proteins regulate cyclic adenosine monophosphate (cAMP) in some signaling pathways. cAMP (3′,5′-cAMP) contains a phosphate group bound to the sugar 5′ and 3′ carbons in a cyclic structure.

D. Degradation of cyclic adenosine monophosphate

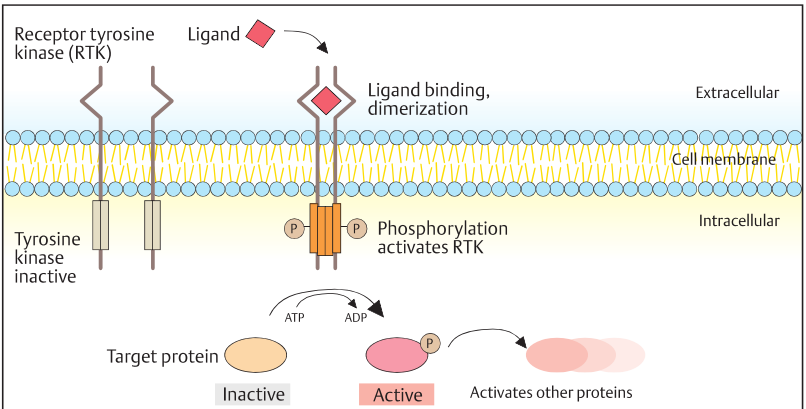
A signal can rapidly increase the cAMP concentration. It is formed by adenylate cyclase. cAMP activates a protein kinase A. Phosphodiesterase rapidly degrades cAMP into inactive AMP. This effect regulates more than 100 signaling proteins.

Medical relevance

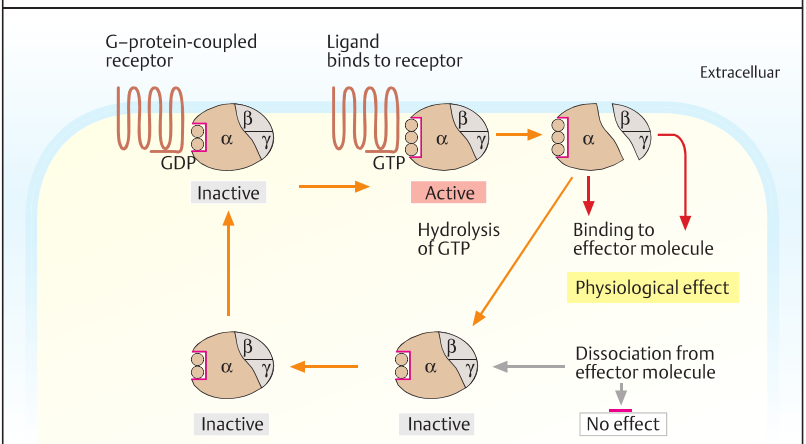
Cholera toxin inhibits GTP hydrolysis and induces a permanently active state. In epithelial cells of the gastrointestinal tract, this causes a massive efflux of water and chloride ions. Pertussis toxin (whooping cough) inhibits adenylate cyclase (inhibitory G protein, G_i) and prevents the α subunit of the inhibitory G protein from interacting with receptors.
Several endocrine disorders result from mutations of genes encoding GPCRs or the G proteins themselves (see Appendix, Table 3, see p. 405).

Further Reading

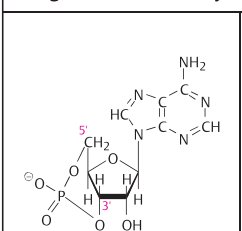
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015
Clapham DE. Mutations in G protein-linked receptors: novel insights on disease. Cell 1993; 75(7): 1237–1239
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016
Newley SE, van Aelst I. Guanine nucleotide-binding proteins. In: Epstein CJ, et al., eds. Inborn Errors of Development. The Molecular Basis of Clinical Disorders of Morphogenesis. 2nd ed. Oxford: Oxford University Press, 2008: 1258–1276



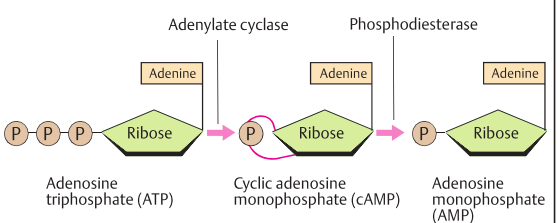
A. Signal transduction by a receptor tyrosine kinase (RTK)



B. Signal transduction by a G–protein-coupled receptor



C. Cyclic AMP



D. Degradation of cAMP

TGF β and Wnt/β-Catenin Signaling Pathways

Transforming growth factor β (TGFβ) and wingless (Wnt)/β-catenin signal transduction pathways are important examples of signaling systems involved in a variety of developmental processes and cellular functions.

A. The Transforming growth factor β signaling pathway

The TGFβ superfamily includes more than 30 structurally related secreted dimeric polypeptides that regulate different biological processes involving growth and differentiation. Binding to a dimeric transmembrane protein as a receptor activates a heterodimeric complex consisting of two copies each of receptors type II and type I (only one copy of each is shown). A constitutively phosphorylated receptor type II phosphorylates the glycogen synthase (GS) domain and activates the receptor type I. As a result, members of the Smad family of transcription factors become serine-phosphorylated and form a complex with a nonphosphorylated common Smad (Co-Smad). Smad 4 moves into the nucleus and binds with DNA-binding proteins to initiate transcription (Smad-dependent, canonical signaling). Non-Smad signaling pathways (noncanonical) may involve the Ras-MAPK (mitogen-activated protein kinase) pathway (see p. 252).

The TGFβ superfamily has important roles in embryonic development, differentiation, organogenesis, self-renewal and maintenance of stem cells in their undifferentiated stage, selection of cell differentiation lineage, and suppression of carcinogenesis. TGFβ is one of the oldest signaling pathways, having arisen approximately 1.3 billion years ago, before the divergence of arthropods and vertebrates (see p. 26). (Figure redrawn from Descotte et al, 2008).

B. Wnt/β-catenin signaling pathway

Wnt are cysteine-rich extracellular signaling glycoproteins regulating cell fate, growth, and differentiation processes. Wingless (*wg*) is a *Drosophila* segment-polarity mutant (see p. 196). The vertebrate ortholog is *Wnt* (derived from wingless and *int-1*, a mouse gene). It regulates β-catenin (in *Drosophila* called *armadillo*) which interacts with other proteins to activate target genes. The receptor for Wnt is a seven-transmembrane receptor named *frizzled* with a coreceptor LRP6, low-density lipoprotein (LDL; see p. 284) receptor related protein. Binding of Wnt to frizzled activates a protein called *dishevelled* (Dsh). Dsh inhibits a protein kinase, GS kinase-3 (GSK3). In the absence of a Wnt signal, β-catenin is phosphorylated by GSK3, APC, and axin (a scaffolding protein) and then degraded (not shown). When frizzled is activated by Wnt binding, unphosphorylated β-catenin accumulates and migrates into the nucleus, binds transcription factors lymphoid enhancer factor, LEF, TCF, (T-cell factor) and others, and regulates Wnt target genes. Of the at least three Wnt-activated pathways downstream of Dsh, the main (canonical) pathway is shown here. Different Wnt signals have roles in development, for example, for gastrulation, brain development, limb patterning, and organogenesis. (Figure adapted from Alberts et al, 2015).

Medical relevance

Dysregulated TGFβ signaling occurs in carcinogenesis and as a cause of Marfan syndrome (see p. 342). Mutations in APC occur in 80% of human cancers of the colon (see p. 320). *WNT4* (OMIM 603490) is a potential ovary-determining gene with mutations resulting disorders of mullerian duct development.

Further Reading

Cadigan KM, Waterman ML. TCF/LEFs and Wnt signaling in the nucleus. Cold Spring Harb Perspect Biol 2012: 4(11): a007906

Descotte V, et al. An introduction to the transforming growth factor β(TGF-β) signaling pathway. In: Epstein CJ, et al., eds. Inborn Errors of Development. The Molecular Basis of Clinical Disorders of Morphogenesis. 2nd ed. Oxford: Oxford University Press, 2008: 358–368

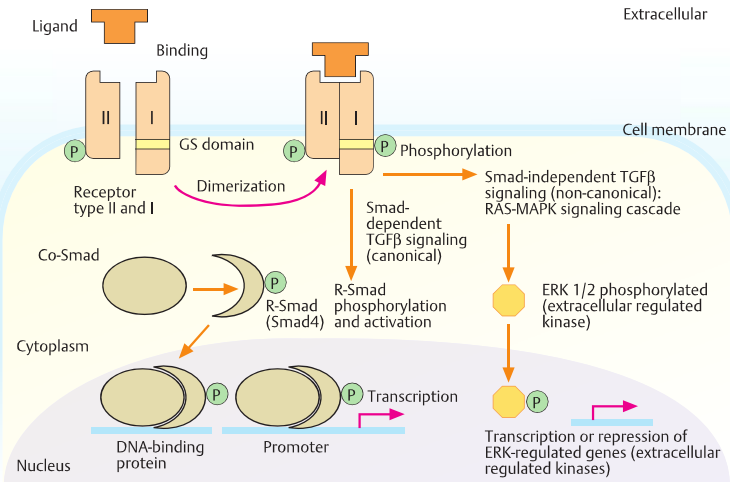
Massagué J. TGFβ signalling in context. Nat Rev Mol Cell Biol 2012; 13(10): 616–630

Moon RT, et al. WNT and β-catenin signalling: diseases and therapies. Nat Rev Genet 2004; 5(9): 691–701

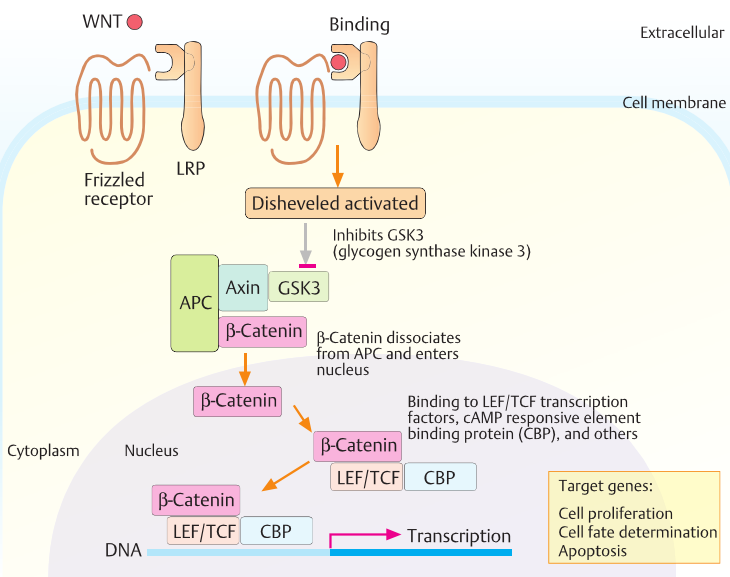
Moses HL, et al. The discovery and early days of TGF-β: a historical perspective. Cold Spring Harb Perspect Biol 2016; 8(7): a021865

OMIM-Online Mendelian Inheritance of Man. Available at www.ncbi.nlm.nih.gov/omim

Sheldahl LC, et al. The Wnt (wingless-type) signaling pathway. In: Epstein CJ, et al., eds. Inborn Errors of Development. The Molecular Basis of Clinical Disorders of Morphogenesis. 2nd ed. Oxford: Oxford University Press, 2008: 330–335



A. TGF β signaling pathway



B. The Wnt/β-catenin signaling pathway

Hedgehog and TNF Signal Pathways

A family of secreted proteins (paracrine factors) called *hedgehog* conveys signals required in many developmental processes, in particular vertebrate limb and neural differentiation. The gene *hedgehog* (*hh*) in *Drosophila* is a segment polarity gene (see p. 196) encoding a transcription factor with both activating and repressing functions in concert with other proteins. Mutant *hh* flies resemble hedgehogs owing to visible spikes on their surfaces. The genomes of vertebrates contain three hedgehog genes: *sonic, desert*, and *Indian hedgehog*.
Tumor necrosis factors (TNFs; OMIM 191160) are macrophage-secreted cytokines with cytotoxic effects. TNFα is involved in two main signaling pathways: (i) kinase and transcription factor activation pathway and (ii) caspase activation in the cell death (apoptosis) pathway (see p. 100). TNFα belongs to a large family of 29 TNF receptors (TNFRs) and 18 TNF ligands in humans.

A. The hedgehog signaling pathway

More than 12 genes form a network. The hedgehog (Hh) signal, secreted as a 45-kDa precursor protein, is the ligand for a transmembrane receptor protein named *patched* (Ptch, PTCH in humans; OMIM 601309). Patched responds to ligand binding by negative control on another transmembrane protein called *smoothened* (Smo, SMOH in humans; OMIM 601500). Smoothened, a protein with seven hydrophobic membrane-spanning domains, acts as a hedgehog signal transducer. Without the signal, a microtubule-bound group of proteins (Costal 2, a kinesin-like protein, and Fused, a serine-threonine kinase) is attached to a 155-kDa effector protein Ci. Two other proteins, PKA (a protein kinase) and Slimb, cleave Ci into two fragments, one of which represses transcription of hedgehog-responsive genes. Binding of the Hh signal to Ptch suppresses the inhibitory function of Ptch on Smo. This inhibits the action of PKA and Slimb. As a result, Costal 2 and Fused are phosphorylated and they activate Ci. Activated Ci moves into the nucleus and initiates transcription together with coactivator cyclic adenosine monophosphate binding protein (CREB/CBP) and other factors.

B. Tumor necrosis factor-α signal transduction

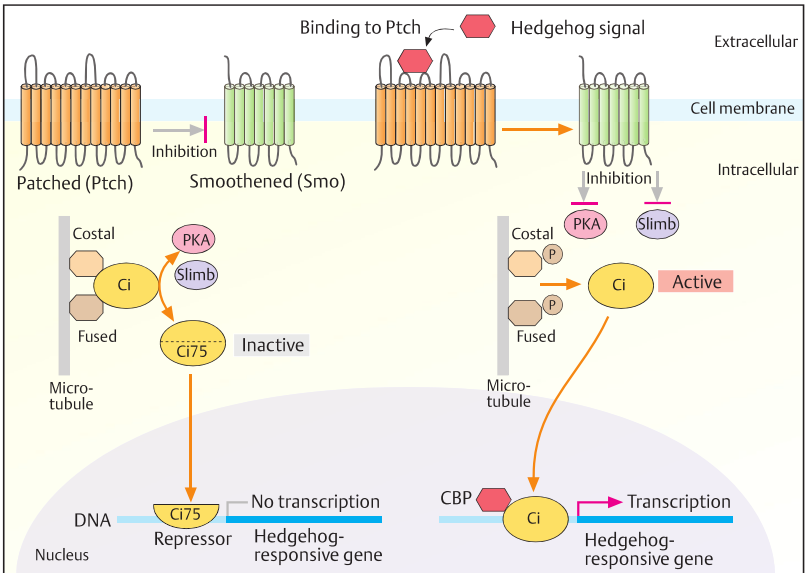
The TNF ligands and their receptors are trimeric proteins. The extracellular moiety of TNFRα consists of two to eight similar structural motifs of elongated shape stabilized by disulfide bridges. Binding of TNFα to the receptor activates the transcription factor NF-κB (nuclear factor kappa B). This causes the recruitment of an intracellular signal protein (receptor-interacting protein kinase), adaptor proteins, TRAF2 (TNF-receptor-associated factor 2), and a death-domain protein (TRADD). (Figures adapted from Gilbert, 2010, and Alberts et al, 2015).

Medical relevance

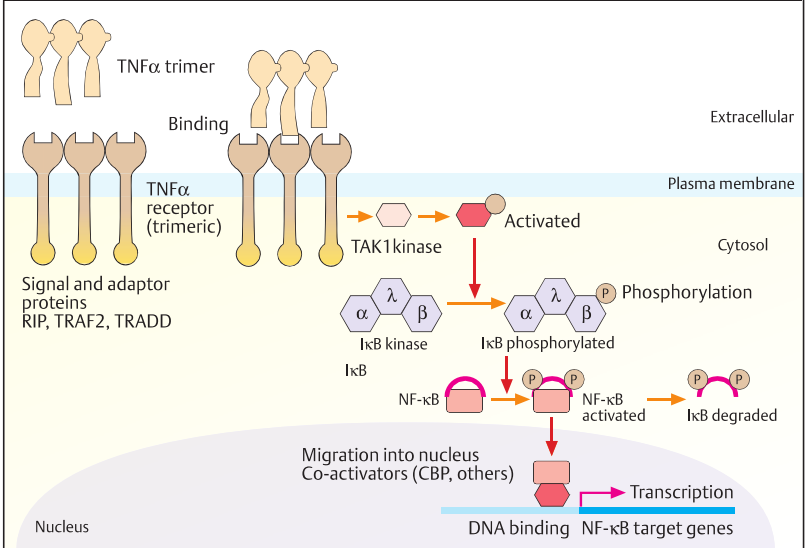
Mutations and deletions in more than 10 human genes in the hedgehog gene network result in a group of malformation syndromes (see Appendix, Table 4, p. 406). Mutations and polymorphisms in TNFα are associated with numerous disorders, for example, rheumatoid arthritis, asthma, colitis ulcerosa, and others.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015
Gilbert SF. Developmental Biology. 9th ed. Sunderland, MA: Sinauer, 2010
Cohen MM. The hedgehog signaling network. Chapter 27. In: Epstein's Inborn Errors of Development. 3rd ed. Erickson RP, et al., eds. Oxford University Press, 2016
Karin M. Nuclear factor-kappaB in cancer development and progression. Nature 2006; 441(7092): 431–436
Lum L, Beachy PA. The Hedgehog response network: sensors, switches, and routers. Science 2004; 304 (5678): 1755–1759
OMIM-Online Mendelian Inheritance of Man. Available at www.ncbi.nlm.nih.gov/omim
Schneider PL. Signaling by TNF and related ligands. Chapter 58. In: Epstein's Inborn Errors of Development. 3rd ed. Erickson RP, et al., eds. Oxford University Press, 2016



A. The hedgehog (Hh) signaling pathway



B. TNFα signaling pathway

The Notch/Delta Signaling Pathway

Notch signaling determines cell specification (e.g., neurogenesis, myogenesis, hematopoiesis), embryo patterning, and morphogenesis of various tissues in the developing vertebrate and invertebrate embryo. Notch signals are transmitted by direct interaction of a sending and a receiving cell. A receptor, Notch, and different types of ligands, called the *DSL group* (Delta, Serrate, Lag-2 ligands), mediate the signal. Notch activation involves regulated proteolytic cleavage at three sites of the receptor.

A. Transmission of a Notch signal

Both the ligand and the receptor are single-pass transmembrane proteins and both require proteolytic cleavage (not shown for the ligand in the schematic figure). The first proteolytic cleavage occurs in the *trans* Golgi network by a furin-like convertase. Notch is O-glycosylated by a glycosyltranferase, fringe, which adds a single fucose to some serines, threonines, and hydroxylysines. Binding to its ligand induces the next two proteolytic cleavages. The Notch tail migrates to the nucleus and binds to the major effector of Notch signaling: the CSL protein. Binding with other gene regulatory proteins induces transcription of Notch-responsive target genes.

B. Lateral inhibition by Notch signaling

A characteristic function of Notch signaling is lateral inhibition in nerve cell development of *Drosophila*. Nerve cells arise as isolated single cells within a sheet of epithelial precursor cells. Cells determined to develop into nerve cells use the Notch signal pathways to inhibit the same cell fate in their neighboring cells. If this fails, a fatal excess of nerve cells causes embryonic death.

C. The *Drosophila* mutant Notch

Notch is a mutant phenotype in *Drosophila melanogaster*, described by T. H. Morgan in 1917. Female heterozygotes exhibit a notch of varied size in the distal part of their wings. Hemizygous males die during embryonic development from hypertrophy of the nervous system.

D. Notch family of receptors

Notch receptors constitute a large family of 50 surface proteins in mammals. The prototype is the *Drosophila* Notch (dNotch), consisting of 36 epidermal growth factor receptor (EGFR)-like tandem repeats in the extracellular domain. Other features are three cysteine (cys)-rich repeats, LNR (Lin-12/Notch-related region), and a signal peptide (SP) at the N-terminal extracellular part. The intracellular domain contains six ankyrin repeats (ANK), flanked by two nuclear localization signals (NLS). Humans have four slightly different types of Notch receptors: hNotch 1–4 (the smaller *Caenorhabditis elegans* notch receptors Lin-2 and Glp-1 are not shown).

E. DSL family of Notch ligands

The Notch ligands likewise comprise a large family of cell surface proteins classified either as Delta or Serrate ligands. They contain multiple EGFR-like repeats in the extracellular domains, an N-terminal DSL-binding motif, and an SP.

Medical relevance

A wide variety of human diseases result from disrupted Notch signaling, for example, Alagille syndrome (OMIM 118450), which is caused by mutations in the gene encoding jagged 1 or *NOTCH2*; spondylocostal dysplasia (OMIM 122600, 271529), resulting from mutations in the *delta-like 3* (*DLL3*) gene; and aortic valve disease (OMIM 109730), resulting from mutations in *NOTCH1* (OMIM 190198).

Further Reading

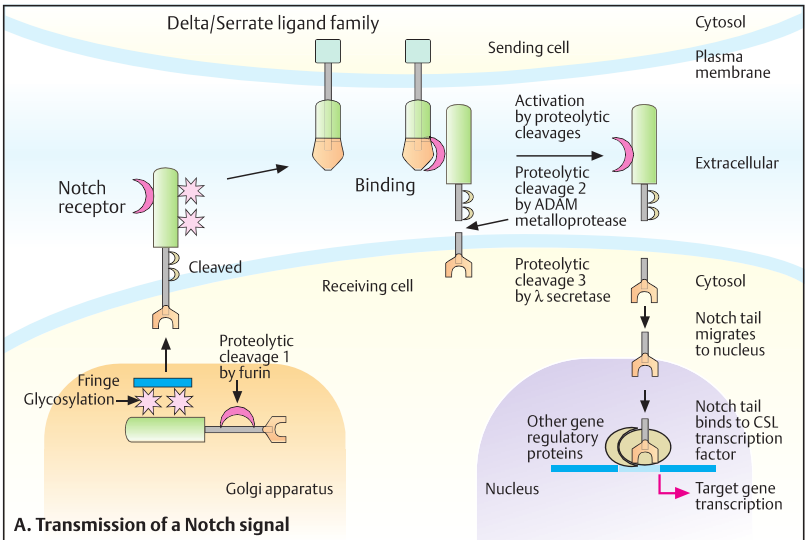
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Krantz ID, et al. Spectrum and frequency of jagged1 (JAG1) mutations in Alagille syndrome patients and their families. Am J Hum Genet 1998; 62(6): 1361–1369

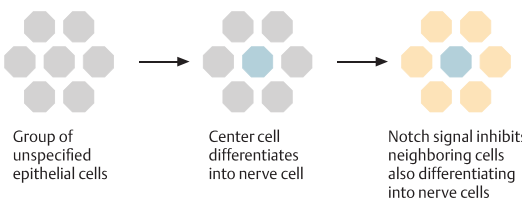
Luca VC, et al. Structural biology. Structural basis for Notch1 engagement of Delta-like 4. Science 2015; 347(6224): 847–853

Miyamoto A, Weinmaster G. The notch signaling pathway. In: Epstein's Inborn Errors of Development. 3rd ed. Erickson RP, et al., eds. Oxford University Press, 2016

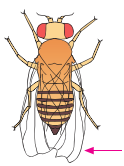
OMIM-Online Mendelian Inheritance of Man. Available at www.ncbi.nlm.nih.gov/Omim



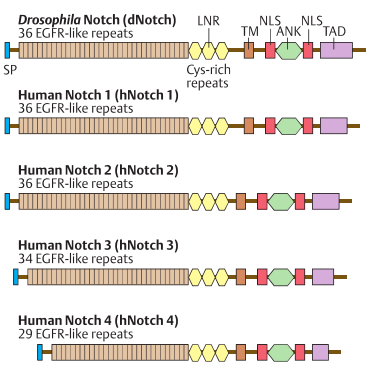
A. Transmission of a Notch signal



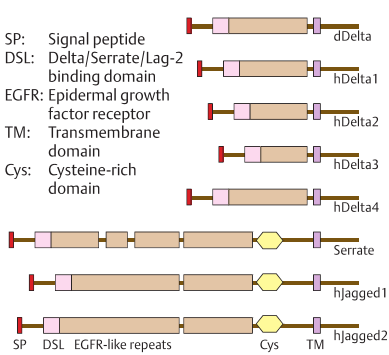
B. Lateral inhibition by Notch/Delta signal in nerve cell development in *Drosophila*



C. The Notch *Drosophila* mutant



D. Notch family of receptors



E. DSL family of Notch ligands

Embryonic Development Genes in *Drosophila melanogaster*

A morphological pattern underlies the development of multicellular organisms. In the *Drosophila* embryo, different groups of genes are expressed at defined stages of development and determine a segmental body plan of the growing embryo. Two main axes are determined: anterior–posterior orientation and dorsal–ventral polarity. Then the segmental pattern of the embryo is established, and, finally, the head, the thorax with legs and antennae, and the abdomen. Homologous genes direct the developmental pattern of the mammalian embryo.

A. Life cycle of the fruit fly

The development from a fertilized egg (~0.5 × 0.15 mm) to the adult fly (2 mm) takes 9 days. Nine nuclear divisions every 8 minutes without cell division leave the nuclei in a syncytium. Following the ninth nuclear division, 90 minutes after fertilization, the nuclei migrate to the periphery and form a *syncytial blastoderm*. After another four nuclear divisions, plasma membranes grow from the periphery and enclose each nucleus. This establishes the *cellular blastoderm* of approximately 6,000 cells. Up to this stage, the embryo depends largely on maternal mRNA and proteins present before fertilization. The embryo passes through three defined larval stages before forming a pupa cocoon. After 5 days of metamorphosis, an adult fly emerges.

B. Segmental organization

The adult fly is organized into 14 segments: three segments (C1–3) form the head, three form the thorax (T1–3), and eight (A1–8) form the abdomen. Each segment has an anterior and a posterior compartment. Each of the initially formed 14 parasegments consists of the posterior compartment of the preceding segment and the anterior compartment of the following segment.

C. Embryonic developmental genes

Three classes of developmental genes are in hierarchical order of development: (i) *egg polarity* maternal effect genes define anterior–posterior and dorsal–ventral polarity in the egg and early embryo; (ii) genes determining the cell fate induce segmentation by *gap* genes and *pair-rule* genes; and (iii) genes determining the structure of body parts after the segmental boundaries are established (*homeotic selector* genes, see next page). The normal embryo shows the region for the head, three segments for the thorax, and eight for the abdomen (**1**). The mutation bicoid in an egg-polarity (maternal effect) gene results in lack of anterior parts (**2**). Mutations in one of approximately nine gap genes induce irregular segmentation named *Krüppel* (**3**) and Knirps (**4**). Characteristic mutant phenotypes of the eight pair-rule genes are even-skipped (**5**) and fushi tarazu (**6**) with loss of portions of alternate segments (fushi tarazu is Japanese for too few segments). More than 10 segment polarity genes determine the anterior–posterior polarity of each segment, for example, the mutant gooseberry (**7**). Homeotic selector genes (**8**) determine the ultimate fate of each segment. In the mutant antennapedia (Ant), the antenna is replaced by a leg (homeotic leg). Many mutations have colorful names derived from the appearance of the mutant embryo and are used untranslated in the language of the discoverers. (Figures adapted from Lawrence, 1992; Nüsslein-Volhard & Wischaus, 1980).

Medical relevance

Mutations in several homologous developmental genes in humans cause disorders with congenital malformations, for example, Waardenburg syndrome type 1 (OMIM 193500).

Further Reading

Albert B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2013

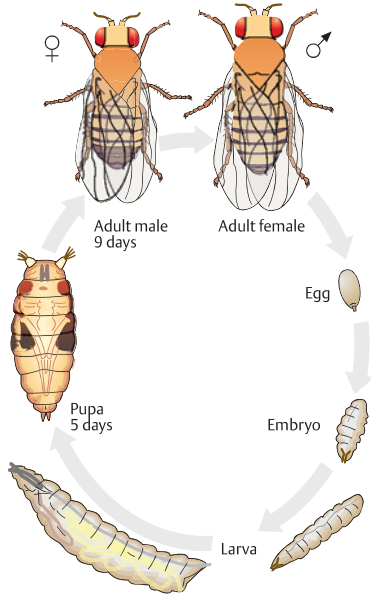
Gilbert SF. Developmental Biology. 9th ed. Sunderland, MA: Sinauer, 2010

Lawrence PA. The Making of a Fly. The Genetics of Animal Design. Oxford: Blackwell Scientific, 1992

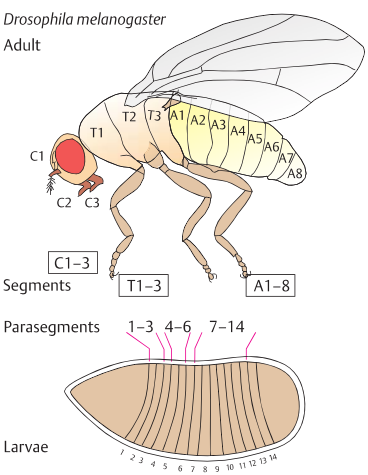
Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in Drosophila. Nature 1980; 287(5785): 795–801

Nüsslein-Volhard C, Frohnhöfer HG. Determination of anteroposterior polarity in Drosophila. Science 1987; 238(4834): 1675–1681

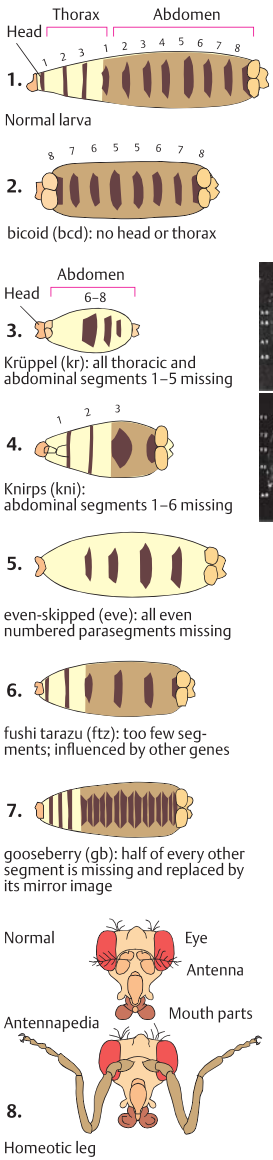
OMIM-Online Mendelian Inheritance of Man. Available at www.ncbi.nlm.nih.gov/Omim



A. Life cycle of *Drosophila melanogaster*



B. Segmental organization of the fruit fly



Gene class

Wild-type

Anterior maternal effect

Pair-rule

Segment polarity

Homeotic genes

C. Embryonic developmental genes

Hox Genes

Hox genes (OMIM 142950ff) form a group of evolutionary related genes that determine basic structures during embryonic development, in particular segmental patterns and their orientation. Originally, they were named homeotic selector genes, now Hox. The named is derived from homeobox, a 180-nucleotide conserved sequence with important functions during embryonic development in *Drosophila* and higher organisms.

A. Regulatory hierarchy of pattern-determining genes

The five principal pattern-determining gene systems in *Drosophila* development act in hierarchical order: (i) *egg polarity* genes, (ii) *gap* genes, (iii) the *pair-rule* genes, (iv) *segment polarity* genes, and (v) *homeotic selector* (*Hox*) genes. Vertebrate Hox genes are homologous to homeotic *Drosophila* genes.

The anterior and posterior system (bicoid for anterior, nanos for posterior) is generated from localized mRNA molecules of maternal origin. The proteins diffuse in a decreasing gradient from the anterior and the posterior pole, respectively. The dorsoventral axis is determined by a transmembrane receptor called *toll*. A fourth signal is generated at both ends, a transmembrane tyrosine kinase receptor called *torso*. The three most important gap genes, *Krüppel*, *Hunchback*, and *Knirps*, determine local patterns. Gap genes induce the pair-rule genes, such as *even-skipped* and *fushi tarazu*. The segment polarity genes, expressed in parasegments, determine the correct anterior–posterior orientation of each individual segment. Homeotic selector genes determine the development of antennae, wings, legs, and other structures.

B. Homeotic selector genes

Homeotic selector genes form a complex called *Hox genes* (HOX in man). *Drosophila* has one set of Hox genes, mammals have four. Each gene belongs to the antennapedia cluster or the bithorax/ultrabithorax (*btx/ubx*) cluster. Five genes of the Antp complex are labial (*lb*), proboscis (*pb*), deformed (*dfd*), sex comb reduced (*scr*), and antennapedia (*antp*). The mammalian Hox genes share a single ancestral homeoselector gene complex. During the evolution of mammals, this was duplicated twice, resulting in 39 genes occurring in four clusters, A–D, in man and mouse. In the process, some genes were lost, while a few genes were added. HOX genes are responsible for pattern formation along the body axis, and are involved in mammalian limb development. They are sequentially expressed in anterior–posterior orientation. (Figure adapted from Alberts et al, 2015).

C. The bithorax mutation

Mutations in the bithorax complex (*Bx-t*) induce the development of an additional thoracic segment with completely developed wings described by Calvin Bridges in 1915. In 1978, E. B. Lewis recognized that the bithorax genes have evolved from a small number of ancestral genes by duplication and subsequent specialization into specific functions.

D. The homeobox

Homeotic selector (*Hox*) genes encode gene regulatory proteins, for example, transcription factors, and are highly conserved in evolution. The antennapedia gene contains a 180-base-pair conserved nucleotide sequence, the homeobox, from which the term *Hox* is derived. In the protein, the corresponding 60 conserved amino acids constitute the homeodomain with four DNA-binding domains (I–IV), a common motif in transcription factors. Their expression depends on segment polarity genes.

Medical relevance

At least 20 human *HOX* genes are involved in various disorders.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

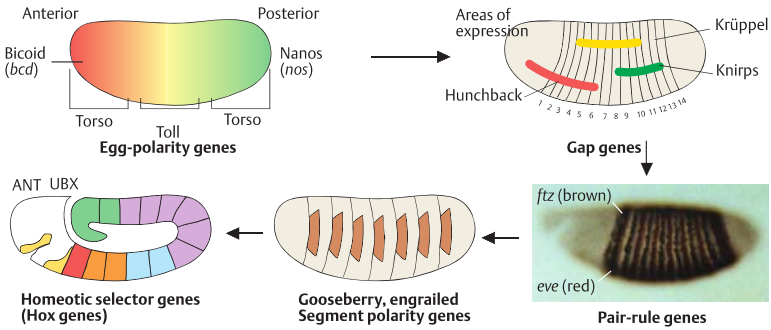
Garcia-Fernàndez J. The genesis and evolution of homeobox gene clusters. Nat Rev Genet 2005; 6(12): 881–892

Krumlauf R. Hox genes in vertebrate development. Cell 1994; 78(2): 191–201

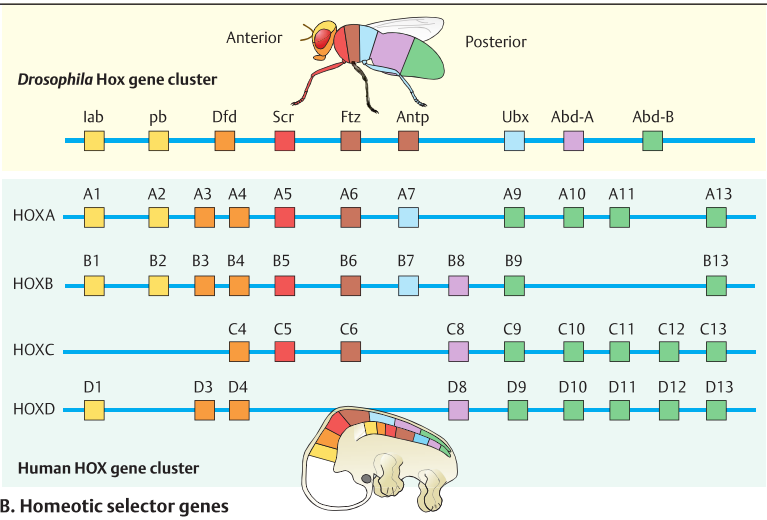
Lawrence PA. The Making of a Fly. The Genetics of Animal Design. Oxford: Blackwell Scientific, 1992

OMIM-Online Mendelian Inheritance of Man. Available at www.ncbi.nlm.nih.gov/Omim

Pearson JC, et al. Modulating Hox gene functions during animal body patterning. Nat Rev Genet 2005; 6(12): 893–904

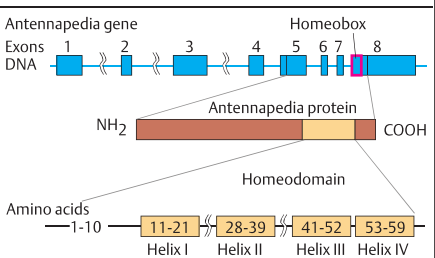


A. Regulatory hierarchy of pattern-determining genes



B. Homeotic selector genes

C. Bithorax mutation



D. Structure of the antennapedia gene and homeobox

Zebrafish: A Translucent Vertebrate

Zebrafish (*Danio rerio*), a tropical freshwater fish, is an important model organism for genetic and genomic studies. It is the first vertebrate to be studied by systematic genetic and genomic analysis. The compact zebrafish genome with approximately 1.5 billion base pairs, half the size of human or mouse, contains 18,000 known protein-coding genes with many known mutations. Owing to a generation time of approximately 3 months and translucent appearance during the first 2 weeks, the effects of many genes in development can be observed.

A. Embryo stages

In the optically clear embryo (pharyngula period), the main parts of the brain (forebrain, midbrain, and hindbrain), and the neural tube, somites, and floor plate are discernible 29 hours after fertilization. At 48 hours (hatching period), pigmentation begins, and the fins, eyes, brain, heart, and other structures become visible. At 5 days (swimming larva), the outline of a fish becomes apparent.

B. Induced mutagenesis

Random mutations induced in male parents were studied in thousands of offspring at various embryonic stages (genetic screen). The mutagenized males were crossed with wild-type females (P), resulting in the first generation (F_1) being heterozygous for mutations (m). Breeding the next generation (F_2) resulted in 50% carrying at least one mutation. Random matings of parents heterozygous for the same mutation resulted in 25% homozygous mutant offspring. Two examples are described in the following.

C. Skeletal phenotype of the fused somites mutation

In wild-type fish, the somite anlagen result in the formation of a normal segmental pattern of the vertebrae and muscles of the trunk and the tail (**1**). Several mutants with abnormal somite boundaries have been identified. Some show posterior somite defects and neuronal hyperplasia, whereas another mutant, called *fused somites* (fss), completely lacks somite formation (**2**). Irregularly shaped spines grow ectopically at the wrong sites. The *fss* gene encodes a T-box transcription factor, Tbx24 (OMIM 607044), required for presomitic mesoderm maturation.

D. The no isthmus midbrain mutation

This mutation, in the *noi* gene (no isthmus), is an example of the numerous distinct mutant phenotypes affecting the central nervous system and spinal cord in zebrafish. Mutant *noi* embryos lack a conspicuous constriction at the boundary between midbrain and hindbrain. Normal 28-hour embryos (wild-type, **1**) show strong expression of the segment polarity gene *engrailed* (*eng*) between the midbrain and hindbrain. Normal eight-somite-stage embryos double stained for *eng* and *krox20* RNA, a marker for rhombomeres 3 and 5 in this region, express *eng* and *krox20* at the midbrain–hindbrain boundary (**2–3**), whereas *noi* mutants do not express *eng* (**4**). The *noi* mutation also eliminates production of wingless (Wnt1) protein in the posterior tectum at the border of the two brain regions in the 20-somite-stage embryo (**5, 6**). (Figures adapted from Brand et al, 1996; van Eeden et al, 1996; Haffter et al, 1996).

Medical relevance

Many human genes carrying disease-causing mutations have a homolog in the zebrafish genome.

Further Reading

Brand M, et al. Mutations in zebrafish genes affecting the formation of the boundary between midbrain and hindbrain. Development 1996; 123: 179–190

Bugel SM, et al. Zebrafish: a marvel of high-throughput biology for 21(st) century toxicology. Curr Environ Health Rep 2014; 1(4): 341–352

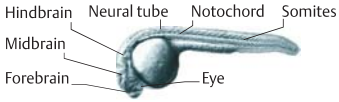
Dodd A, et al. Zebrafish: bridging the gap between development and disease. Hum Mol Genet 2000; 9 (16): 2443–2449

Haffter P, et al. The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. Development 1996; 123: 1–36

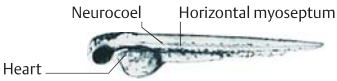
Howe K, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature 2013; 496: 498–503

van Eeden FJ, et al. Mutations affecting somite formation and patterning in the zebrafish, Danio rerio. Development 1996; 123: 153–164

ZFIN: The Zebrafish Information Network. Available at http://zfin.org/. Accessed January 3, 2017



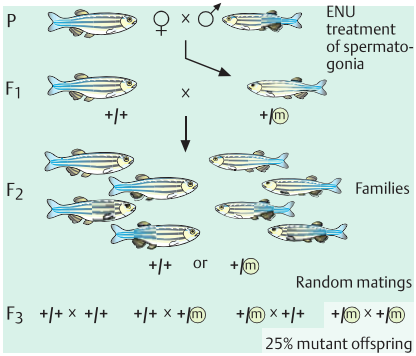
1. 29 hours, pharyngula period



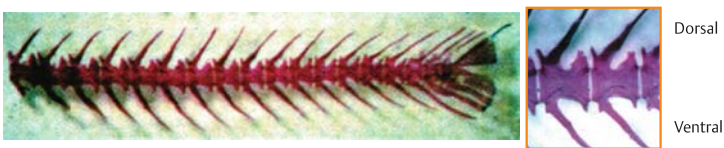
2. 48 hours, hatching period

3. 5 days, swimming larva

A. Embryo stages



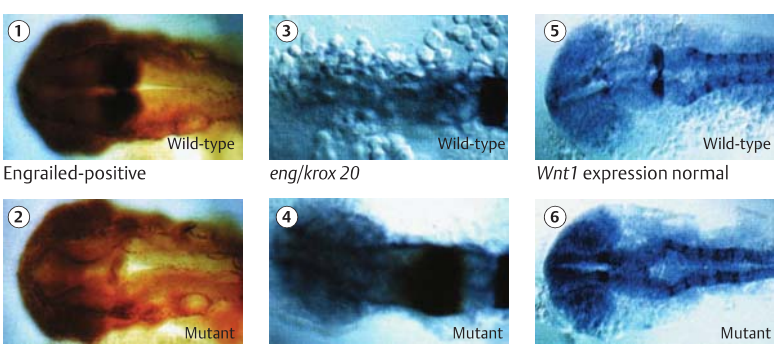
B. Induced mutagenesis and genetic screen



1. Wild-type

2. *fss* mutant

C. Skeletal phenotype of fused somites (*fss*) mutations



Engrailed-positive | *eng/krox 20* | *Wnt1* expression normal

noi-negative | *eng*-negative | *Wnt1* expression abolished in MHB

D. The midbrain mutation no isthmus (*noi*)

Cell Lineage in a Nematode, *Caenorhabditis elegans*

Caenorhabditis elegans is a free-living transparent worm consisting of a precise number of somatic cells, each of which can be traced back to a founder cell. Sydney Brenner introduced this tiny worm as a model organism in 1965. Systematic genetic analysis of many mutant phenotypes has yielded important insights into the interaction of genetic, anatomical, and physiological traits in development. The complete cell lineage has been established in this organism.

The 97-Mb genome of *C. elegans* contains approximately 20,000 protein-coding genes and more than 160,000 genes encoding untranslated RNA. Approximately 32% of the coding sequences are homologous to sequences in man, and approximately 70% of known human proteins have homologies in *C. elegans*. The largest group of genes encodes transmembrane receptors (790), in particular chemoreceptors, zinc finger transcription factors (480), and proteins with protein-kinase domains. Interfering RNA (see p. 174) has important functions in *C. elegans*.

A. Caenorhabditis elegans

Caenorhabditis elegans is a transparent worm, approximately 1 mm in length, with a life cycle of approximately 3 days. Its basic structure is a bilaterally symmetric elongated body of nerves, muscles, skin, intestines, and gonads. It exists as one of two sexes: hermaphrodite or male. Hermaphrodites produce eggs, and spermatozoa and can reproduce by self-fertilization. The adult hermaphrodite worm has 959 somatic cell nuclei; the adult male worm has 1,031. In addition, there are 1,000 to 2,000 germ cells.

B. Origin of individual cells

All tissues arise from six founder cells, AB for the nervous system, MS for muscle cells, E for intestine, C for skin, D for some muscles, and P4 for germ cells. At each cell division, genetically established rules determine the fate of each of the two daughter cells. Differentiated cells are derived from more than one founder cell, except for cells of the intestine and gonad. Of the 959 adult cells, 302 are nerve cells.

C. Developmental control genes

Many genes directing development have been identified by analysis of the mutations induced by ethylmethanesulfonate. The principal types of mutation may cause a cell to differentiate into a wrong cell type (e.g., Z instead of B), or to divide too early or too late (division mutants).

D. Apoptosis in C. elegans

Programmed cell death (apoptosis) is a normal part of vertebrate and invertebrate development (see p. 110). During embryonic development of *C. elegans*, 131 of 947 nongonadal cells of the adult hermaphrodite undergo apoptosis at a defined time and branching point during different stages of development (**1**). Mutations in the *ced-9* gene induce apoptosis. Normally, *ced-9* suppresses apoptosis, and ced-3 and *ced-4* are proapoptotic genes. Apoptosis does not occur in *ced-9*/*ced-3* double mutants because *ced-9* is upstream of *ced-3* in the apoptosis pathway. The photographs (**2**) show the death of a cell (cell P11.aap) over a time span of approximately 40 minutes. (Figures adapted from Wood, 1988, and Sulston & Horvitz, 1977).

Medical relevance

The human *BCL-2* gene, a homolog of *ced-9*, encodes an inner mitochondrial membrane protein that inhibits apoptosis in pro-B lymphocytes. Disruption of this gene causes follicular lymphoma, a B-cell tumor (OMIM 151430).

Further Reading

Brenner S. The genetics of Caenorhabditis elegans. Genetics 1974; 77(1): 71–94

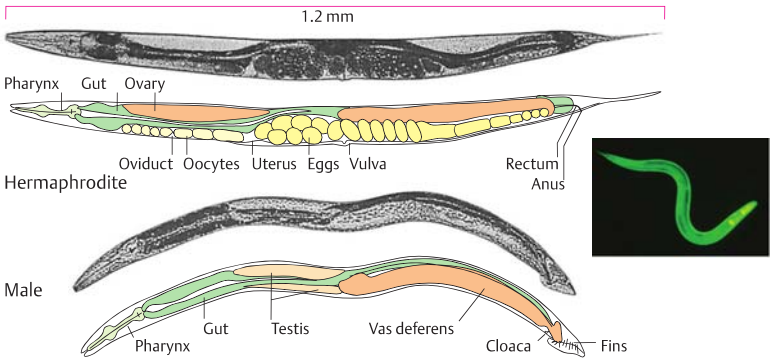
C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 1998; 282(5396): 2012–2018

Culetto E, Sattelle DB. A role for Caenorhabditis elegans in understanding the function and interactions of human disease genes. Hum Mol Genet 2000; 9(6): 869–877

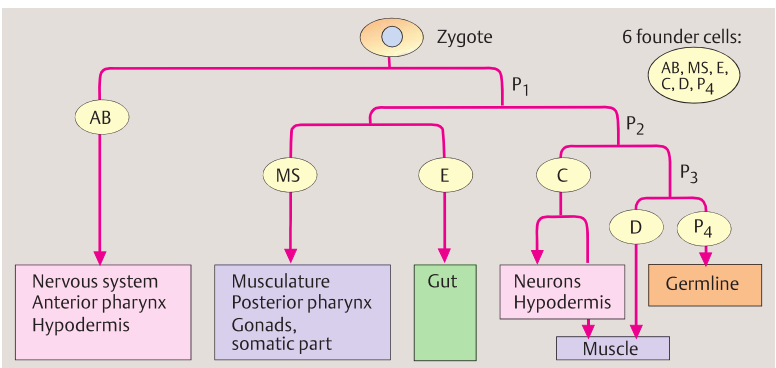
Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev Biol 1977; 56(1): 110–156

Wood WB, ed. The Nematode *Caenorhabditis elegans*. Monograph 17. New York, NY: Cold Spring Harbor Laboratory Press, 1988

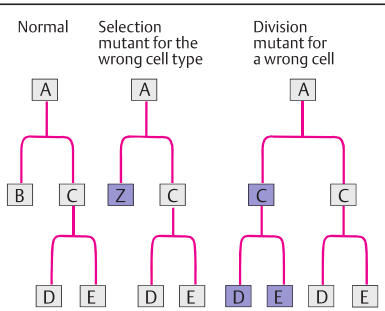
WormBase. Available at http://www.wormbase.org/. Accessed July 24, 2017



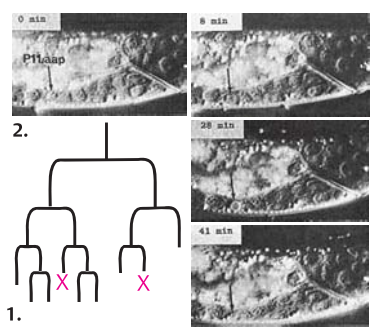
A. *Caenorhabditis elegans*



B. Embryonic origin of individual cells



C. Mutations in developmental control genes



D. Programmed cell death (X)

Genomics

Genomics: The Study of the Organization of Genomes

The genome represents all molecules and functional elements required for survival and genetic transmission to the next generation. This term can apply to a single cell, an individual member of a species, or an entire species. Genomics is the scientific field dealing with all structural and functional aspects of the genome. Increasingly efficient automated sequencing methods have yielded the DNA sequence of the genome of numerous animals, thousands of humans from different ethnic groups, plants, and all microorganisms of medical or agricultural interest.

The term genomics includes all molecules involved in transcription and translation and their regulation (the *transcriptome*); all approximately 90,000 proteins that a cell or an organism is able to produce (the *proteome*); the functional analysis of all genes (*functional genomics*); and the assembly, storage, and management of data (*bioinformatics*), and any other relevant area.

A. Examples of organisms whose genomes have been sequenced

This plate illustrates the great diversity of organisms whose genome has been sequenced during the past 25 years. The DNA sequence of the human genome has been compared with that of an extinct hominid lineage, the Neandertal and others, and that of the great apes, other primates, and most mammals and marsupials.

Sources of images: **1**, A. Dürer, 1507. Adam und Eva. Museo Nacional del Prado, Madrid; **2**, J. Weissenbach, 2004; **3**, Weissenbach J. Genome sequencing: differences with the relatives. Nature 2004; 429: 353–355; **4**, The Jackson Laboratory; **5**, Robert Geisler, Max Planck Institute for Developmental Biology, Tübingen, Germany; available at http://zf-health.org/; **6**, Marco van Kerkhoven; **7**, M. Ashburner, University of Cambridge, United Kingdom; **8**, Dr. Jordan Boyle; **9**, Wikimedia Commons; **10**, Escherichia coli, Wikipedia; **11**, The Arabidopsis Information Resource (TAIR).

Early sequencing results

Homo sapiens: International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004; 431: 931–945

Lander ES. Initial impact of the sequencing of the human genome. Nature 2011; 470: 187–197.

Nature: Human Genome Collection. Available at http://www.nature.com/nature/supplements/collections/humangenome/

UCSC Genome Bioinformatics. Available at http://genome.ucsc.edu/. Accessed February 2, 2017

Skipper M. User's guide to the human genome. Nature Rev Genet 2012; 13: 678

Neandertal: Green RE, Krause J, Briggs AW, et al. A draft sequence of the Neandertal genome. Science 2010; 328: 710–722

Chimpanzee: Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 2005; 437: 69–87

Dog: Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 2005; 438: 803–819

Mouse: Waterston RH, Lindblad-Toh K, Birney E, et al.; Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature 2002; 420: 520–562

Rat (not shown): Gibbs RA, Weinstock GM, Metzker ML, et al.; Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 2004; 428: 493–521

Zebrafish (*Brachydanio rerio*): Sanger Institute. Available at http://www.sanger.ac.uk/resources/zebrafish/. Accessed January 27, 2012

Anopheles gambiae, the vector of the malaria parasite *Plasmodium falciparum:* Holt RA, Subramanian GM, Halpern A, et al. The genome sequence of the malaria mosquito Anopheles gambiae. Science 2002; 298: 129–149

***Drosophila melanogaster*:** Adams MD, Celniker SE, Holt RA, et al. The genome sequence of *Drosophila melanogaster*. Science 2000; 287: 2185–2195

***C. elegans*:** C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 1998; 282: 2012–2018

Yeast: Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. Science 1996; 274: 546, 563–567

Bacterium *Escherichia coli*: Blattner FR, Plunkett G 3rd, Bloch CA, et al. The complete genome sequence of Escherichia coli K-12. Science 1997; 277: 1453–1462

Plant: Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 2000; 408: 796–815

Sea Urchin: Sea Urchin Genome Squencing Consortium. The genome of the sea urchin Strongylocentrotus purpuratus. Science 2006; 314: 941–952

1. *Homo sapiens*
3000 Mb ~22 000 genes

2. Chimpanzee
3000 Mb
1.2% differences
to humans

3. *Canis domesticus*
24100 Mb, 19 300 genes

4. *Mus musculus*
2500 Mb ~25 000 genes

5. Zebrafish 1600 Mb
(70% sequenced) ~22 000 genes

6. *Anopheles gambiae*
278 Mb, ~14 000 genes

7. *Drosophila melanogaster*
180 Mb, 13 600 genes

8. *C. elegans*
97 Mb, 19 000 genes

9. *S. cerevisiae*
12.1 Mb, 6300 genes

10. *E. coli* 4700 kb
4300 genes

11. *Arabidopsis thaliana*
125 Mb, ~ 25 000 genes

A. Examples of organisms whose genomes have been sequenced

Genomes of Microorganisms

Bacterial genomes are small, ranging from approximately 130 kilobases (kb) to 14 million base pairs, and are tightly packed with genes aligned nearly contiguously along a circular chromosome. Most bacteria contain approximately 1,000 to 6,000 genes. The coding regions are small (average 1 kb) and contain no introns. Genome sequences from tens of thousands of bacterial genomes from 50 different bacterial phyla and 11 different archaeal phyla are available. As of February 2015, approximately 14,000 bacterial and archaeal genome sequences have been determined.

Owing to their small size, bacterial genomes were the first to be completely sequenced, for example, *Haemophilus influenzae*, in 1995. The smallest bacterial genome is that of *Mycoplasma genitalium* (483 genes). It is an obligate intracellular pathogen. It lacks genes encoding proteins for metabolic functions. The limited capacity for metabolism in *M. genitalium* is compensated by the transport of life-supporting molecules from its extracellular environment into the cell.

A. The genome of a small bacteriophage

Phage ΦX174, with 10 genes (A–J) contained in 5,386 nucleotides in single-stranded DNA, was the very first organism to have its genome sequenced (Sanger et al., 1977). The genome of phage ΦX174 is so compact that several genes overlap.

B. Overlapping genes in ΦX174

The reading frames of genes A and B, B and C, and D and E partially overlap. The overlapping genes are transcribed in different reading frames: in the start codon ATG of gene E (above the sequences shown), the last two nucleotides (AT) are parts of the codon TAT for tyrosine (Tyr) in gene D. Similarly, the stop codon TGA for the E gene is part of codons GTG (valine) and ATG (methionine) in the D gene. Thus, this small genome is used very efficiently.

C. Genome of *Escherichia coli*

This simplified figure shows the essential features of a bacterial genome. Functionally related genes usually cluster in operons (four of many shown). Approximately half of the genes of are in operons. In general, the size of a bacterial genome corresponds to the number of genes. Bacteria existing in a symbiotic environment have the smallest genomes.

Further Reading

Bentley SD, et al. Comparative genomic structure of prokaryotes. Annu Rev Genet 2004; 38: 771–792

Blattner FR, et al. The complete genome sequence of Escherichia coli K-12. Science 1997; 277(5331): 1453–1462

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Fleischmann RD, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995; 269(5223): 496–512

NCBI. Microbial Genomes. Available at www.ncbi.nlm.nih.gov/genome/microbes/. Accessed July 24, 2017

Land M, et al. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 2015; 15 (2): 141–161

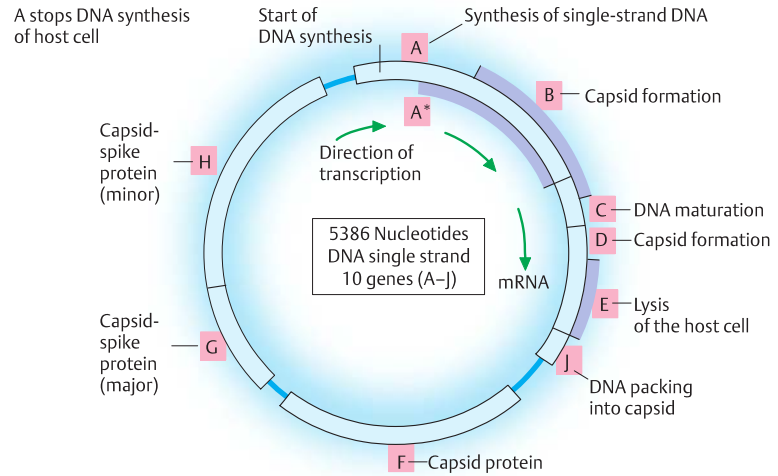
Microbial Genomics. Bases to Biology.

Ochman H, Davalos LM. The nature and dynamics of bacterial genomes. Science 2006; 311(5768): 1730–1733

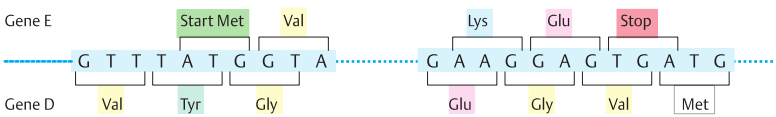
Size and general contents of bacterial genomes

Main genomic features	Free-living	Facultative pathogen	Pathogen or obligate symbiont
Genome size	Large (5–10 Mb)	2–5 Mb	Small (0.5–1.5 Mb)
Genome stability	Stable or unstable	Unstable	Stable
Lateral gene transfer	Frequent	Frequent/rare	Rare or none
Number of pseudogenes	Few	Many	Rare
Population size	Large	Small	Small
Pathogenicity factors	Absent	Present	Present

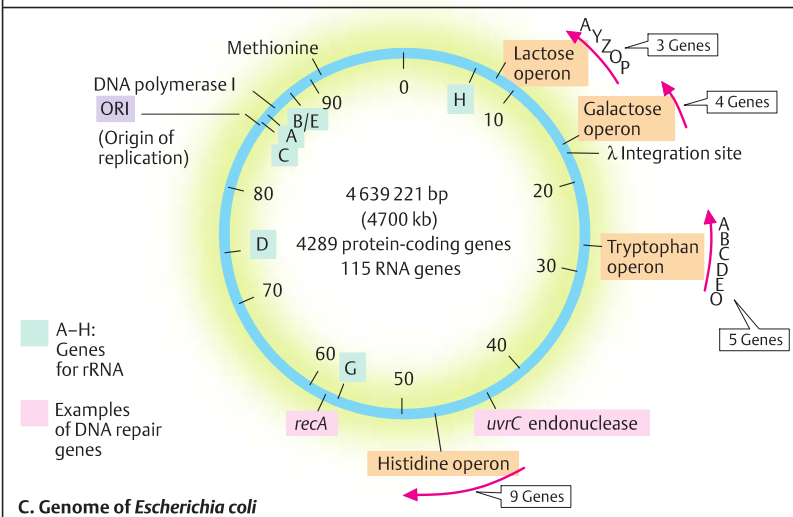
Abbreviation: Mb, million base pairs. Note: Data from Ochman and Davalos (2006).



A. Genome of the bacteriophage ΦX174



B. Overlapping genes in ΦX174



C. Genome of *Escherichia coli*

Architecture of the Human Genome

The human genome contains 2.85 billion (2.85 Gb) nucleotides (approximately 3×10^9 base pairs [bp]) per haploid genome. If each nucleotide is printed as a letter of 1 mm width, this would correspond to a chain of approximately 3,200 km in length. The approximately 21,000 protein-coding genes make up approximately 1.5% of the genome. Approximately 6% of the human DNA sequence consists of conserved noncoding elements. Basic features of the human genome are a low average density of genes along a given stretch of DNA; duplicated chromosomal segments; interspersed, repetitive sequences; and evidence of transposition events. The repeats derived from transposons arose during evolution mostly through RNA intermediates and make up approximately 40% of the genome.

A. Types of sequences

The types of sequences can be classified into sequences that are parts of genes or their regulatory parts (gene-related) and repetitive sequences outside genes (intergenic sequences). These consist of different types of interspersed repeat sequences.

B. Interspersed repetitive DNA

Genome-wide repeats consist of mainly four types that are present at many sites. *Long interspersed repeat sequences* (LINEs, **1**) are mammalian retrotransposons that lack long terminal repeats in contrast to retroviruses. LINEs consist of repetitive sequences up to 6,500 bp long. LINEs encode two open reading frames (ORF1 and ORF2), which are translated. In addition to a 5′ promoter (P), they have an internal promoter. Approximately 600,000 L1 elements are dispersed throughout the human genome. LINE-2 and LINE-3 are inactive because reverse transcription from the 3′ end often fails to proceed to the 5′ end. LINE-1 accounts for 21% of the genome.

Short interspersed repeat sequences (SINEs, **2**) are repetitive segments of approximately 100 to 400 bp with tandem duplication of CG-rich segments separated by A-rich segments. They do not encode a protein and are not capable of autonomous insertion. The most abundant type of SINEs in the primate genome is the Alu family (Alu sequences) with approximately 1.2 million copies in humans (approximately 6% of the genome). One Alu repeat occurs about once every 3 kilobases (kb) in the human genome.

Long terminal direct repeat (*LTR*) *retroposons* (**3**) are flanked by LTRs containing transcriptional regulatory elements. The autonomous retrotransposons contain *gag* and *pol* genes, which encode proteins required for retrotransposition.

DNA transposons (**4**) form different classes resembling bacterial transposons with inverted repeats and encoding a transposase.

C. Segmental duplications

The human genome contains more than 1,000 blocks of duplicated segments, which account for 5% of the genome. They are located mainly in the subtelomeric and pericentromeric regions or occur interspersed along each chromosome (segmental duplications).

They occur as blocks of 1- to 200-kb genomic sequences that are also present at another site in a chromosome (intrachromosomal) or another chromosome (interchromosomal). The three examples show segmental duplications present in the X chromosome: chromosome 20 and chromosome 4. Blocks shared with other chromosomes are shown as connecting lines. When unequal crossing-over occurs in this region a DNA stretch with a duplication or a deficiency may result, leading to a *genomic disorder* (see p. 236).

Medical relevance

LINE and SINE movements can lead to sporadic disease (e.g., hemophilia A, see p. 268).

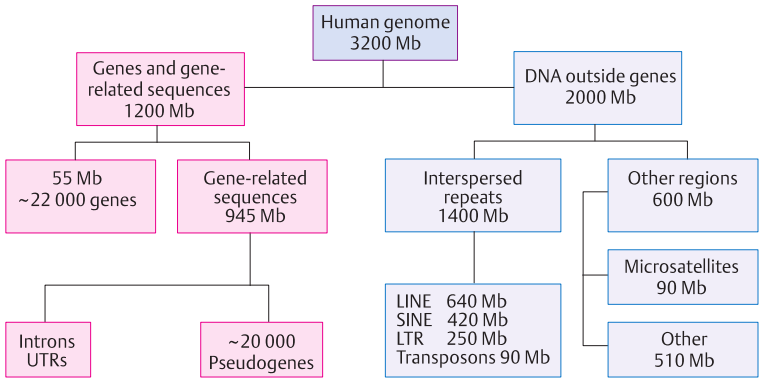
Further Reading

Beck CR, et al. LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet 2011; 12: 187–215

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Emanuel BS, et al. Segmental duplications: an 'expanding' role in genomic instability and disease. Nat Rev Genet 2001; 2(10): 791–800

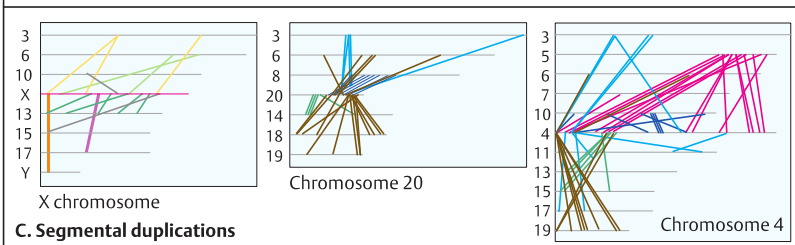
Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015



A. The types of sequence in the nuclear human genome

1. Long interspersed elements (LINE) (autonomous)	P ORF 1 ORF 2 $(A)_n$ 6–8 kb	LINE-1 ~600 000 copies LINE-2 ~370 000 copies LINE-3 ~44 000 copies
2. Short interspersed elements (SINE) (non-autonomous)	$(A)_n$ 100–300 bp	Alu family ~1200 000 copies MIR ~450 000 copies MIR3 ~85 000 copies
3. Retrovirus-like elements (LTR transposons)	LTR gag pol (env) LTR P 6–11 kb	Human endogenous retroviral sequences (HERV) ~240 000 copies (several classes; autonomous and non-autonomous)
4. DNA transposons	Transposase 2–3 kb	Several classes (autonomous and non-autonomous) ~300 000 copies

B. Major types of interspersed repetitive DNA in the human genome



C. Segmental duplications

Regulatory Architecture of the Human Genome

The tightly packaged DNA in chromatin must be accessible to transcription to allow gene expression in different cell types and developmental pathways. The genome functions in a hierarchical structure at different levels involving a closely connected variety of functional regulatory elements. These are contained in distinct physical domains called *topologically associated domains* (TADs). Here regulatory control is exerted by *cis*-regulatory elements such as enhancers, silencers, and insulators, often within locus control regions. Through the formation of loops, they come into contact with promoters of their target genes. Active chromatin relates to other active regions, whereas repressed chromatin relates to other silent regions. The ENCODE (Encyclopedia of DNA Elements) project systematically maps regions of transcription, transcription factor associations, chromatin structure, and histone modifications. Its results have allowed functional aspects to be assigned to approximately 80% of the genome.

A. Hierarchical organization of the genome

The DNA (**1**) packaged into nucleosomes (**2**) forms a primary structure. This is partitioned into distinct TADs (**3**) as part of a chromosome (**4**), corresponding to a secondary structure. The overall chromosome folding (**5**) represents a tertiary structure. This model is reminiscent of the folding structure in proteins.

B. Topologically associated domains

A TAD consists of evolutionarily highly conserved sequences with an average size of 1 million base pair (**1**). Within a nuclear TAD, enhancer and promoter interact over a distance of several 100 kilobases (DNA) by forming a loop (**2**). The borders between TADs in a genomic region containing isolated binding sites (**3**) are formed by an isolator element. This prevents loop formation and proximity of the binding sites (**4**). The boundary separating domains (TDBs, topological domain boundaries) is associated with a transcriptional repressor CTCF, which links to cohesion and forms a loop. CTCF is a highly conserved 11-zinc finger protein encoded by the *CTFC* gene (OMIM 604167). It is involved in regulatory functions, including transcriptional activation/repression, insulation, imprinting, and X chromosome inactivation.

C. Functional states of topologically associated domains

Sexton and Cavalli distinguish facultative and constitutive TAD models. In the facultative model (**A**), repressed (blue) and active (red) domains form separate TAD domains. Some genes may leave the repressed state and are activated during development by a shift in the boundary. In contrast, in the constitutive model (**B**), state boundary positions make no change. Here gene expression occurs by an altered intra-TAD interaction. (Figures adapted from Sexton & Cavalli, 2015).

Medical relevance

Structural rearrangements of enhancer, silencer, or insulator elements lead to tissue-specific loss of function or aberrant gene expression (see Disorders Resulting from Structural Rearrangements of *Cis*-Regulatory Elements, p. 240).

Further Reading

Birney E; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489(7414): 57–74

Cavalli G, Mistelli T. Functional implications of genome topology. Nat Struct Mol Biol 2013; 20(3): 290–299

Dixon JR, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012; 485(7398): 376–380

de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 2013; 502(7472): 499–506

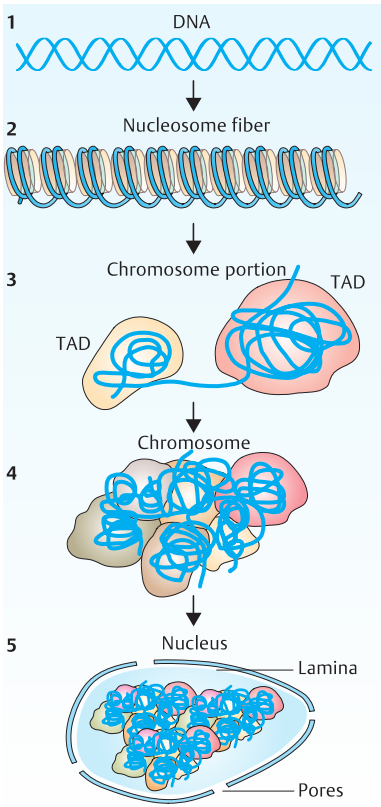
Ibn-Salem J, et al. Deletions of chromosomal regulatory boundaries are associated with congenital disease. Genome Biol 2014; 15(9): 423

Ong CT, et al. CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 2014; 15(4): 234–246

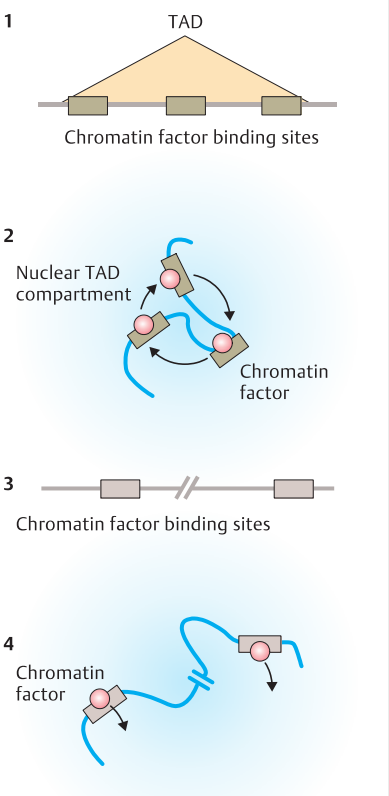
Thurman RE, et al. The accessible chromatin landscape of the human genome. Nature 2012; 489 (7414): 75–82

Sexton T, Cavalli G. The role of chromosome domains in shaping the functional genome. Cell 2015; 160 (6): 1049–1059

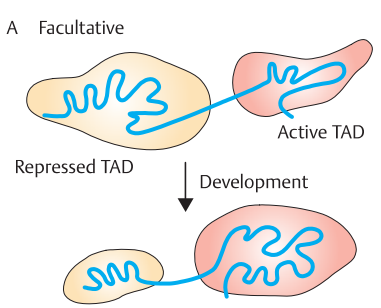
Spielmann M, Mundlos S. Structural variations, the regulatory landscape of the genome and their alteration in human disease. BioEssays 2013; 35(6): 533–543



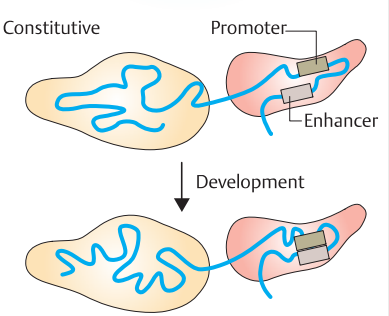
A. Hierarchical organization of the genome



B. Topologically associated domains (TADs)



C. Functional states of TADs



Genome Analysis with DNA Microarrays

The genome can be assessed for diagnostic purposes by a variety of microarrays. A microarray or DNA chip is an assembly of oligonucleotides or other DNA probes fixed on a small, fine grid of surfaces. Such high-density microarrays can be used to analyze complex RNA or DNA samples. With a microarray, the expression states of many genes can be examined simultaneously. One approach is to use complementary DNA (cDNA) prepared from messenger RNA (mRNA; expression screening) or to recognize sequence variations in genes (screening for DNA variation). The advantages of using microarrays are manifold: simultaneous large-scale analysis of thousands of genes at a time, automation, small sample size, and easy handling. Several manufacturers offer highly efficient microarrays that contain more than 6 million different oligonucleotide populations on a 1.7-cm^2 (1.28 × 1.28 cm) surface.
Two basic types of DNA microarrays can be distinguished: (i) microarrays of previously prepared DNA clones or polymerase chain reaction products that are attached to the surface, arranged in a grid of high-density arrays in two-dimensional linear coordinates, and (ii) microarrays of oligonucleotides synthesized in situ on a suitable surface.

A. Gene expression profile by complementary DNA array

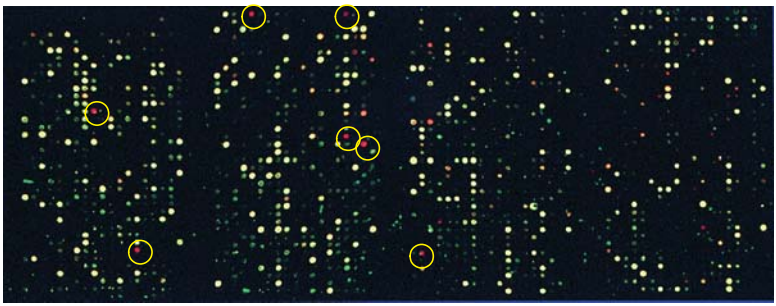
This figure shows a microarray of 1,500 different cDNAs derived from the human X chromosome. The cDNAs were obtained from lymphoblastoid cells of a normal male (XY) and a normal female (XX). The cDNAs of the female cells were labeled with the fluorophore Cy3 (red), and cDNAs of the male cells were labeled with Cy5 (green). The inactivation of most genes in one of the two X chromosomes in female cells leads to a 1:1 ratio of cDNAs from the expressed genes in the male and female X chromosomes. The result is a yellow signal at most sites because the superimposed red fluorescent (female) and green fluorescent (male) signals are present in equal amounts. At seven sites, marked by yellow circles, the signal is red. Genes that have escaped inactivation on the inactive female X chromosome (see p. 184) are expressed at a double dosage. (Photograph provided by Drs. GM Wieczorek, U Nuber and HH Ropers, Max Planck Instititute Molecular Genetics, Berlin).

B. Gene expression patterns in human cancer cell lines

Microarrays can be used to analyze the pattern of gene expression in cancer cells, as an aid to diagnosis and the monitoring of therapy. In this figure, the gene expression patterns in 60 cell lines derived from different human cancers are shown. Approximately 8,000 genes have been analyzed by this procedure (Ross et al., 2000). A consistent relationship between gene expression patterns and tissue of origin was detectable.
A cell line dendrogram is shown, relating the patterns of gene expression to the tissue of origin of the cell lines as derived from 1,161 cDNAs in 64 cell lines (1). Also shown is a colored microarray representation of the data using Cy3-labeled (red) cDNA reverse-transcribed from mRNA isolated from the cell lines compared with Cy5-labeled (green) cDNA derived from reference mRNA (2). Several clusters of red dots in the columns (1,161 genes) and the rows (60 cell lines) indicate an increased gene expression. Thus, these regions of the microarray represent genes with altered gene expression patterns in tumor cells. (Images from Ross et al, 2000, with permission).

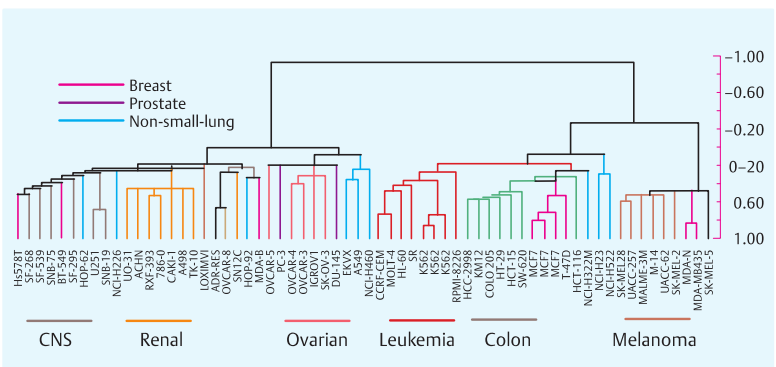
Further Reading

Affymetrix GeneChip. Available at www.vsni.co.uk/software/genstat/htmlhelp/marray/AffymetrixChips.htm
Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007
Gresham D, et al. Comparing whole genomes using DNA microarrays. Nat Rev Genet 2008; 9(4): 291–302
Illumina Science & Education (Technology, Sequencing, Medical Genetics, Publication Reviews): www.illumina.com/science.html. Accessed 25 July, 2017
Ross DT, et al. Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 2000; 24(3): 227–235
Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011
Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015



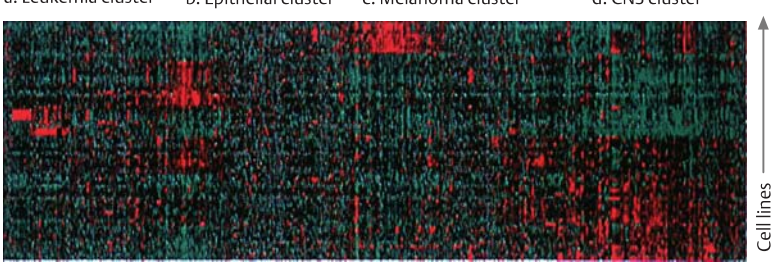
A. Gene expression profile by cDNA array

XIST Expression



1. Cell line dentogram

a. Leukemia cluster b. Epithelial cluster c. Melanoma cluster d. CNS cluster



2. Microarray data

B. Gene expression patterns in human cancer cell lines

Genome Scan and Array–Comparative Genomic Hybridization

A whole-genome scan and array–comparative genomic hybridization (CGH, see next page) aim at the analysis of the entire genome rather than selected groups of genes. Genome-wide studies are now widely used in the analysis of linkage and association (see p. 118). One type of whole genome scan is a single-nucleotide polymorphism (SNP) array. It detects polymorphic DNA sequence variants (SNPs, see p. 72) at multiple loci along all chromosomes in search for association with a region or linkage that might harbor gene loci or chromosomal regions involved in predisposing to a disease or causing it. More than 12 million SNP variants have been identified in the human genome (SNP database) to be used in research or diagnostics. Array-comparative genomic hybridization (aCGH) combines the microarray methodology with genomic hybridization. The principle is illustrated schematically here.

A. Principle of whole-genome scan

A whole-genome scan analyzes all chromosomes (only four are shown here schematically) in search for linkage of a disease locus with polymorphic markers along each chromosome (multipoint linkage analysis). A peak of the logarithm of the odds (to the base 10) (LOD) score of nearly 4 over chromosome 2 indicates that linkage between a disease locus and two marker loci (shown in red and blue) may be present. A LOD score of 3 or above indicates a probability ratio for linkage/against linkage of 1,000:1 or above.

In addition, a genome-wide scan may aim at detecting an association between a region containing susceptibility loci for complex diseases and polymorphic variants. An association will be detected by linkage disequilibrium (LD). LD is the result of a preferential segregation of certain alleles with a susceptibility region (see p. 118). Many different computer-based test procedures for analyzing linkage exist.

B. Array–comparative genomic hybridization

aCGH is a combination of microarray techniques and CGH (see next page). In aCGH, many mapped clones in a microarray are used instead of metaphase chromosomes. This greatly increases the detection rate for small deletions or duplications. The genome to be tested and a normal genome as reference are compared. They are differently labeled. In the example shown here, the DNA from an individual to be tested is labeled with the fluorophore Cy5, which induces green fluorescence. The DNA from individual used as a control is labeled with Cy3, which induces red fluorescence.

If the amount of DNA in the diagnostic test and the control sample is the same, a yellow signal will result. If there is an imbalance, the signal will be red if the amount of DNA is reduced due to a deletion. A green signal will indicate a duplication. aCGH is a fast and efficient means of detecting deletions and duplications on a genome-wide scale. (Figure kindly provided by Dr. Evan E Eichler, Univ. of Washington, Seattle).

Medical relevance

aCGH is widely used in diagnostics.

Further Reading

Database of Genomic Variants. A curated catalogue of human genomic structural variation. Available at http://dgv.tcag.ca/dgv/app/home. Accessed February 6, 2017

Eichler EE. Widening the spectrum of human genetic variation. Nat Genet 2006; 38(1): 9–11

Hinds DA, et al. Whole-genome patterns of common DNA variation in three human populations. Science 2005; 307(5712): 1072–1079

International HapMap Project. Available at https://www.genome.gov/10001688/. Accessed February 6, 2017

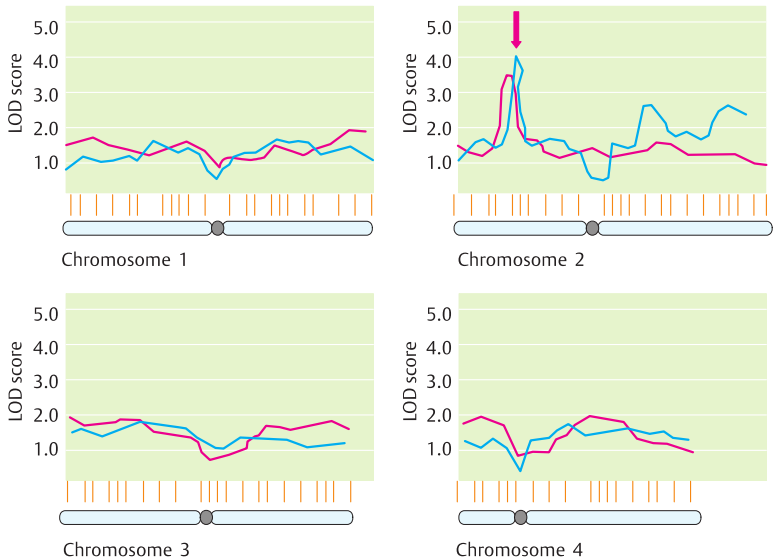
NCBI. SNP Database. Available at https://www.ncbi.nlm.nih.gov/nucleotide/. Accessed February 6, 2017

Pinkel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998; 20(2): 207–211

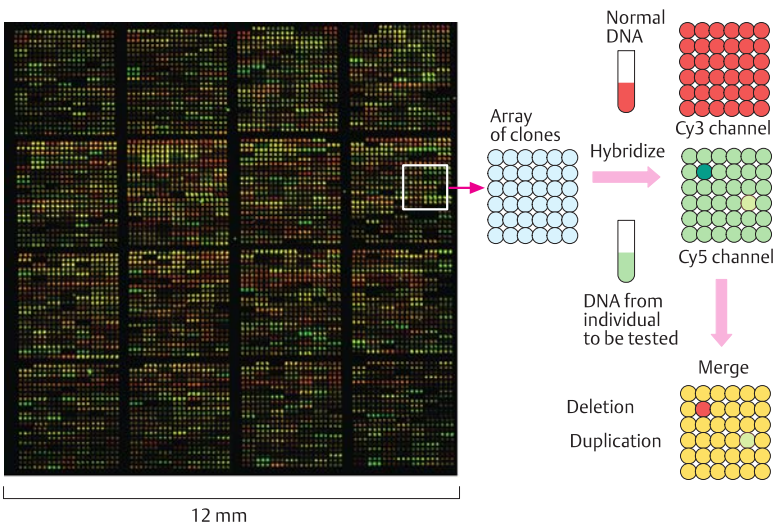
Database SNP. Available at http://www.hgvs.org/central-mutation-snp-databases. Accessed February 6, 2017

Thomas DC, et al. Recent developments in genome-wide association scans: a workshop summary and review. Am J Hum Genet 2005; 77(3): 337–345

Vissers LE, et al. Identification of disease genes by whole genome CGH arrays. Hum Mol Genet 2005; 14 Spec No. 2: R215–R223



A. Principle of whole-genome scan



B. Array-comparative genomic hybridization

Comparative Genomic Hybridization

Comparative genomic hybridization (CGH) represents a systematic search for small deletions and duplications (microdeletions/microduplications) at the level of the whole genome. It detects differences in the copy number variation (CNV, see p. 236) of DNA in all chromosomes simultaneously by comparing DNA from a patient with that of an unaffected control. It combines fluorescence in situ hybridization and chromosome painting by using probes that hybridize to entire chromosomes. CGH is widely used to investigate tumor cells in metaphase or interphase. Modifications of CGH are array-based CGH (aCGH, see the previous page) and multiplex ligation-dependent probe amplification (MLPA; Sellner and Taylor, 2004).

A. Application of CGH

CGH can detect changes in the genome of tumor cells. It compares DNA isolated from a tumor sample with genomic DNA from a normal individual as a control. The figure shows a metaphase (top) from a patient with a hereditary form of renal cell carcinoma (von Hippel–Lindau syndrome; OMIM 193300). On the left, the metaphase is labeled green by fluorescein isothiocyanate, which in this case labels tumor DNA (**1**). In the middle, the same meta-phase is labeled red by rhodamine labeling control DNA (**2**). On the right (**3**) is an electronically blended composite of the fluorophore labels of green and red. Yellow fluorescence indicates a normal 1:1 green:red ratio without loss or gain of chromosomal material. Red fluorescence indicates chromosomal loss in the tumor, whereas green fluorescence indicates chromosomal gain. The hybridization pattern is shown for all chromosomes in the CGH profile (**4**). The individual chromosomes are identified by a blue dye that stains AT-rich sequences (DAPI, 4′,6-diamidino-2-phenylindole), visualized by fluorescence microscopy. The metaphase is sequentially analyzed with a charge-coupled device camera for red, green, and blue (DAPI counterstain) using different filter systems. The blue DAPI fluorescence is converted into a black-and-white image. This is enhanced to result in a banding pattern similar to G-bands for chromosome identification. The CGH profile scans each chromosome for a deviation to the red (first vertical line on the left of each field right to a chromosome) to search for loss, or for a deviation to green (third vertical line), which will indicate a gain. A deviation beyond the red line to the left is visible along the long arm of chromosome 4. This indicates chromosomal loss in 4q. A deviation to the green over the long arm of chromosome 10 indicates a gain of chromosomal material in 10q. The centromeres are shown as horizontal gray boxes. (Images courtesy of Drs. Nicole McNeil and Thomas Ried, National Institutes of Health).

B. Identification of extra chromosomal material by multicolor fluorescence in situ hybridization

Here, extra chromosomal material, invisible in a standard karyotype, is visualized by multicolor fluorescence in situ hybridization (M-FISH, see p. 154). The multiplex FISH karyogram (right) shows a small extra band at the end of the long arm of a chromosome 1 (arrow). This extra band is derived from a chromosome 12. (Photographs provided by Drs. Sabine Uhrig and Michael Speicher, Graz, Austria).

Further Reading

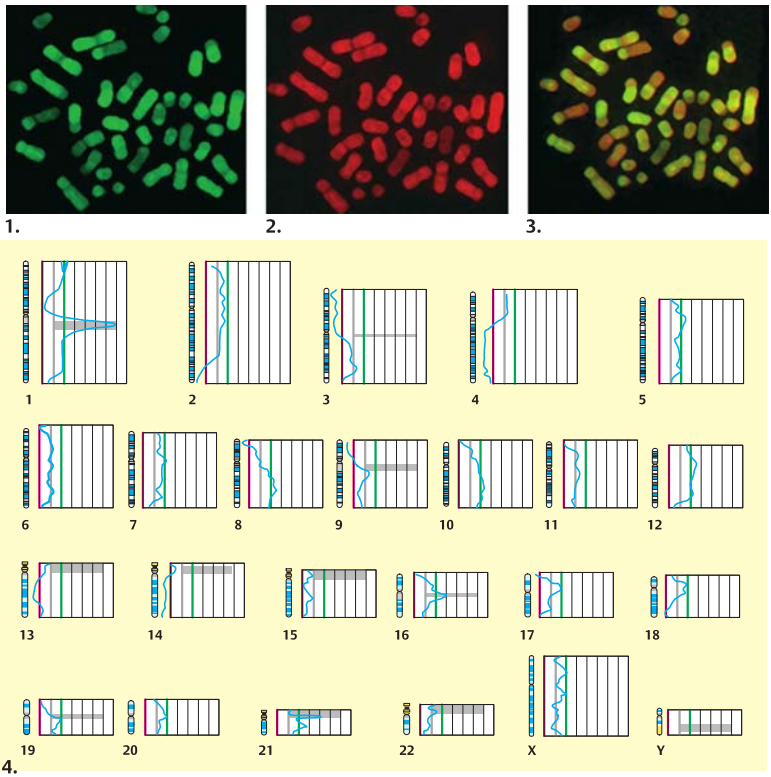
Knutsen T, et al. The interactive online SKY/M-FISH & CGH database and the Entrez cancer chromosomes search database: linkage of chromosomal aberrations with the genome sequence. Genes Chromosomes Cancer 2005; 44(1): 52–64

Sellner LN, Taylor GR. MLPA and MAPH: new techniques for detection of gene deletions. Hum Mutat 2004; 23(5): 413–419

Speicher MR, Carter NP. The new cytogenetics: blurring the boundaries with molecular biology. Nat Rev Genet 2005; 6(10): 782–792

Vissers LE, et al. Identification of disease genes by whole genome CGH arrays. Hum Mol Genet 2005; 14 Spec No. 2: R215–R223

Wong A, et al. Detection and calibration of microdeletions and microduplications by array-based comparative genomic hybridization and its applicability to clinical genetic testing. Genet Med 2005; 7(4): 264–271



A. Comparative genomic hybridization (CGH)

B. Identification of extra chromosomal material by M-FISH

Genome-Wide Association Study

Genome-wide association study (GWAS) is a whole-genome approach to search for chromosomal regions or known gene loci for an association with a disease or risk for disease. Hundreds of thousands of single-nucleotide polymorphisms (SNPs, see p. 72) are tested in large populations of individuals affected and compared with large numbers of controls. More than 2,000 GWASs involving 17 different genetic traits have been performed since 2013. Each SNP has a reference SNP number (rs) recorded in the National Center for Biotechnology Information dbSNP database, with information on allelic forms, allele frequency, and other genomic information. Problems may arise from false-positive results in the huge number of statistical tests performed. GWA genotyping platforms on arrays or chips comprising 500,000 to 1 million SNPs are available.

A. Graphic representation of GWAS data

The association of SNPs with a chromosomal region or known genes is represented in a so-called Manhattan plot. This displays all human chromosomes with a graphic representation of the SNPs determined along each chromosome. It is shown typically as the $-\log_{10}$ of the *P*-value (probability of the observed association arising by chance). Thus, $P = 0.01$ would be plotted as 2 or $P = 10^{-7}$ as 7 on the y-axis. In the example shown, 317,503 SNPs were studied in 1,522 individuals with rheumatoid arthritis and 1,850 controls. In addition to the major histocompatibility (MHC) region (here *HLA-DRb1*) on the short arm of chromosome 6 and the *PTPN22* gene on chromosome 1 known to be related to the immunological aspects of the disease, two additional genes with a possible etiological role were identified: *TRAF1* (encoding tumor necrosis factor receptor associated factor-1) and *C5* (encoding complement component 5).

B. Identification of common variant in a complex disorder

In this example, a GWAS of two cohorts of 780 families (3,101 subjects) and 1,204 affected individuals with an autism spectrum disorder are compared with 6,491 control subjects. Six SNPs between two genes, *CDH10* and *CDH9* (encoding neuronal cell-adhesion molecules cadherin 9 and 10), show a strong association of signals on the short arm of chromosome 5 (5p14.1) with a $-\log_{10}$ *P*-value above 5 (**1**). Chromosome 5 is shown schematically (**2**) below the plot for all chromosomes. In (**3**), the enlarged region 5p14.1, the nucleotides 24,500,000 to 27,000,000, and the two genes and their 5′ to 3′ orientation are shown. SNPs are plotted with their *P*-values (**4**). Genotyped SNPs are colored according to their correlation with one of the SNPs (rs4307059): red for above 0.5, yellow between 0.5 and 0.2, and white for below 0.2. (Figure in A adapted from Plenge et al, 2007; in B from Wang et al, 2009, both with permission from Macmillan Publishers Ltd).

Medical relevance

Examples disease gene or genome region identification by GWAS: Crohn disease, age-related macular degeneration, diabetes mellitus, neuropsychiatric diseases, Alzheimer disease, hypertension, and others.

Further Reading

NIH-National Human Genome Research Institute. A Catalog of Published Genome-Wide Association Studies. Available at https://www.genome.gov/26525384. Accessed July 15, 2017

Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. PNAS 2009; 106(23): 9362–9367

Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med 2010; 363(2): 166–176

Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at www.genome.gov/gwastudies. Accessed July 25, 2017

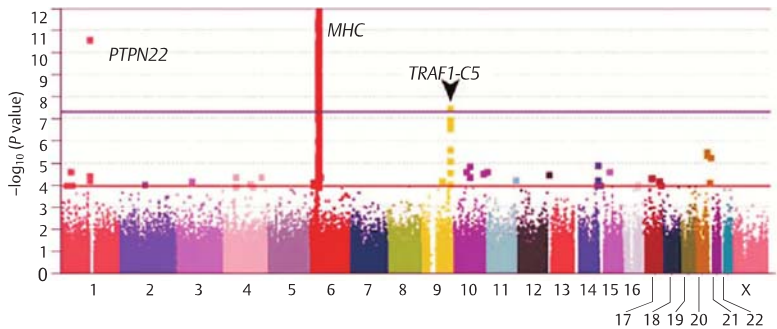
Plenge RM, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. N Engl J Med 2007; 357(12): 1199–1209

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015

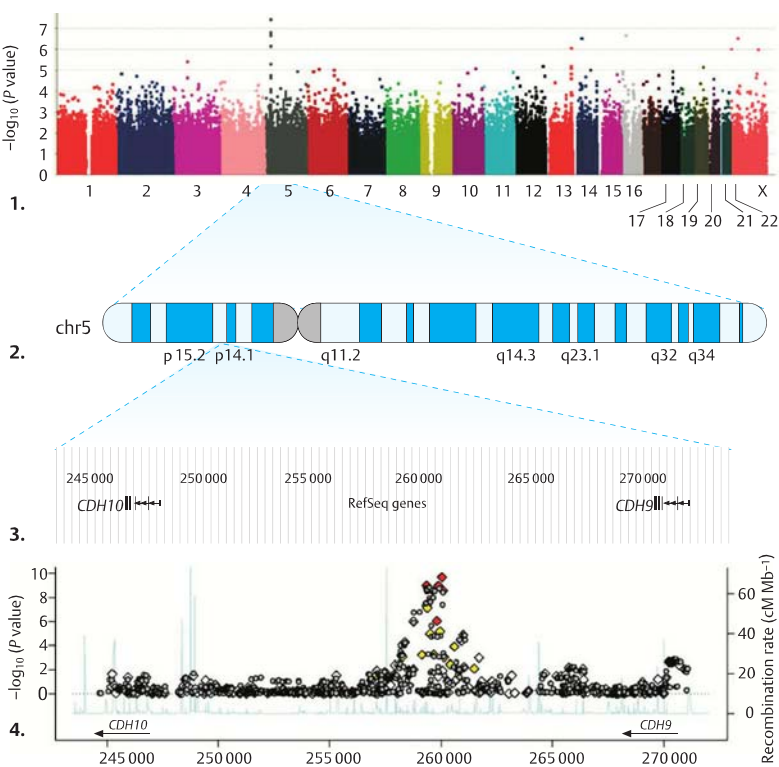
Visscher PM, et al. 10 years GWAS discovery: biology, function, and translation. Am J Hum Genet 2017; 101: 5–22

Wang K, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 2009; 459(7246): 528–533

Welter D, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014; 42(Database issue): D1001–D1006



A. Graphic representation of GWAS data



B. Identification of common variant in a complex disorder

The Dynamic Genome: Mobile Genetic Elements

Mobile genetic elements are DNA sequences that are able to move from one site to another by transposition (see p. 82). The term *dynamic genome* indicates that the genome of all living organisms is not static. Instead, it is flexible and subject to changes. This unusual phenomenon was first observed in the late 1940s by Barbara McClintock while investigating the genetics of Indian corn (maize, *Zea mays*). She found that certain genes were able to alter their position spontaneously. She named this phenomenon "jumping genes," and later *mobile genetic elements*. Although McClintock's observations were initially met with skepticism, she was awarded the Nobel Prize for this work in 1984 (Fox-Keller, 1983; McClintock, 1984). Transposable genetic elements or their remnants account for nearly half of the genome in mammals and up to 90% of the genome in some plants. Transposition provides mechanisms for genomes to acquire new sequences and rearrange existing ones during the course of evolution.

A. Stable and unstable mutations

McClintock in 1953 determined that certain mutations in maize are unstable. A stable mutation at the C locus causes violet corn kernels (**1**), whereas unstable mutations cause fine pigment spots in individual kernels (variegation, **2**).

B. Effect of mutation and transposition

Normally, a gene at the C locus produces a violet pigment of the aleurone in cells of Indian corn (**1**). This gene can be inactivated by insertion of a mobile element (*Ds*) into the gene, resulting in a colorless kernel (**2**). If *Ds* is removed by transposition, C-locus function is restored and small pigmented spots appear (**3**).

C. Insertion and removal of Ds

Activator-dissociation (Ac/Ds) is a system of controlling elements in maize. *Ac* is an inherently unstable autonomous element. It can activate another locus, dissociation *(Ds*), and cause a break in the chromosome (**1**). While *Ac* can move independently (autonomous transposition), *Ds* can move to another location in the chromosome only under the influence of *Ac* (nonautonomous transposition). The Ac locus is a 4.6-kilobase (kb) transposon; Ds is defective without a transposase gene (see p. 82). The C locus (**2**) is inactivated by the insertion of *Ds*. Ds can be removed under the influence of *Ac*. This restores normal function at the C locus.

D. Transposons in bacteria

Transposons are classified according to their effect and molecular structure: simple insertion sequences (IS) and the more complex transposons (Tn). A transposon contains additional genes, for example, for antibiotic resistance in bacteria.

Transposition is a special type of recombination by which a DNA segment of approximately 750 base pairs to 10 kb is able to move from one position to another, either on the same or on another DNA molecule. The insertion occurs at an integration site (**1**) and requires a break (**2**) with subsequent integration (**3**). The sequences on either side of the integrated segment at the integration site are direct repeats. At both ends, each IS element or transposon carries inverted repeats whose lengths and base sequences are characteristic for different IS and Tn elements.

(Photographs in A and B from Fedoroff, 1984).

Medical relevance

Insertion of a transposon into a functional gene may be a rare cause of a genetic disease.

Further Reading

Fedoroff NV. Transposable genetic elements in maize. Sci Am 1984; 250: 65–74

Fedoroff NV, Botstein D, eds. The Dynamic Genome: Barbara McClintock's Ideas in the Century of Genetics. New York, NY: Cold Spring Harbor Laboratory Press, 1992

Fox-Keller E. A Feeling for the Organism: The Life and Work of Barbara McClintock. San Francisco, CA: W. H. Freeman, 1983

Kazazian HH Jr. Mobile DNA: Finding Treasure in Junk. Upper Saddle River, NJ: FT Press Science, Pearson Education, 2011

Kazazian HH Jr. Mobile elements: drivers of genome evolution. Science 2004; 303(5664): 1626–1632

Krebs JE, et al. Lewin's Genes XI. Sudbury: Bartlett & Jones, 2013

McClintock B. Induction of instability at selected loci in maize. Genetics 1953; 38(6): 579–599

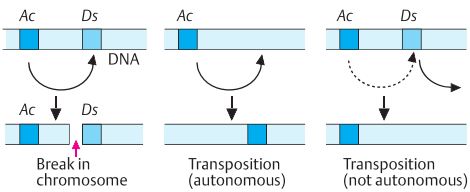
McClintock B. The significance of responses of the genome to challenge. Science 1984; 226(4676): 792–801

1. Stable mutation (violet pigment) **2.** Variegation by unstable mutation (fine spots)

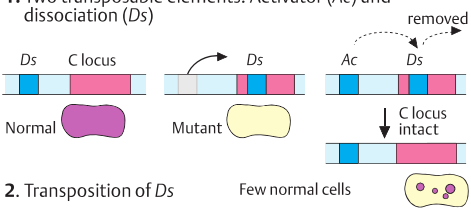
A. Stable and unstable mutations in maize

1. Violet pigment formation at the C locus **2.** Inactivation by transposable element *Ds* **3.** Removal of *Ds* from some cells by *Ac*

B. Effect of mutation and transposition

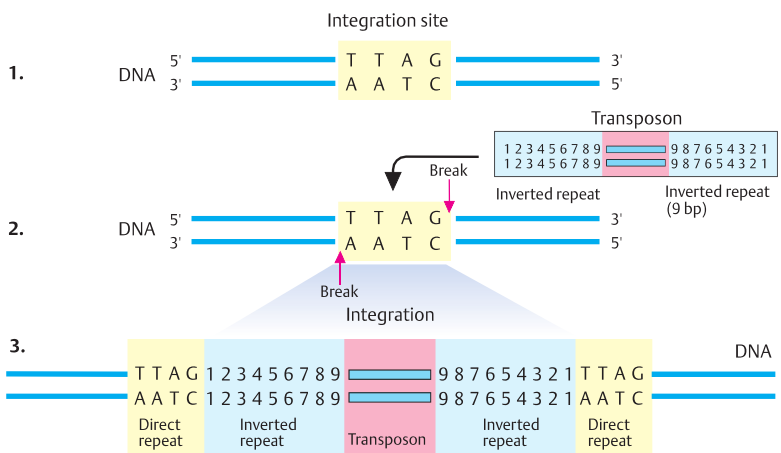


1. Two transposable elements: Activator (*Ac*) and dissociation (*Ds*)



2. Transposition of *Ds*

C. Insertion and removal of *Ds*



D. Transposons in bacteria

Genome Editing by the CRISPR-Cas System

Genome editing refers to an efficient assembly of molecular methods that allow producing precise, targeted changes in the genome. The CRISPR-Cas system (*C*lustered *R*egularly *I*nterspaced *S*hort *P*alindromic *R*epeats and associated protein-9 endonuclease, Cas9) is derived from bacteria where it functions as a defense system against invading foreign DNA. Viral DNA sequences are integrated into a CRISPR locus array of the bacterial genome. Upon reinfection, an adaptive immune system based on CRISPRs identifies invading foreign DNA sequences and induces a site-specific double-strand break by means of CRISPR-associated endonuclease Cas9.

A. Function of a CRISPR-Cas system in bacteria

The CRISPR locus consists of several hundred repeats interspersed with foreign sequences derived from prior exposure (spacers). An adaptive immunity results from two phases: integration of short viral DNA into a specific repeat CRISPR locus (**1**), and, upon reinfection, invading foreign DNA is recognized and cleaved (**2**). This is achieved by transcription into precursor CRISPR RNA, which is cleaved at each repeat into mature crRNAs.

The crRNA forms a duplex with the transactivating crRNA, which is loaded onto Cas9 to direct the cleavage of the target DNA adjacent to PAM sequences (protospacer adjacent motifs, 5'-NGG-3').

B. Use of the CRISPR-Cas system in genome editing

For genome-editing, the crRNA and tracRNA are fused to form a single guide RNA, which hybridizes to one strand of the invading target DNA. Two nuclease domains of Cas9 induce site-specific breaks in each strand of the target DNA. Since DNA breaks induce DNA repair pathways in host cells, this process can be used to disrupt genes (knockout) or replace DNA sequences (knock-in).

Medical applications

The CRISPR/Cas system has potential to be used in gene therapy to correct a mutation and to remove infectious agents such as hepatitis B, HIV, and other applications. (Figure in A redrawn from Alberts et al, 2015; in B from Charpentier & Doudna, 2013).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015: 434 and 497

Barrangou R, Marraffini LA. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol Cell 2014; 54(2): 234–244

Charpentier E, Doudna JA. Biotechnology: rewriting a genome. Nature 2013; 495(7439): 50–51

Chu VT, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol 2015; 33(5): 543–548

Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 2014; 346(6213): 1258096

Doudna JA, Sternberg SH. A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution. Houghton Mifflin Harcourt, New York, 2017

Hsu PD, et al. Development and applications of CRISPR-Cas9 for genome engineering. Cell 2014; 157(6): 1262–1278

Jinek M, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337(6096): 816–821

Lander ES. The heroes of CRISPR. Cell 2016; 164: 18–28

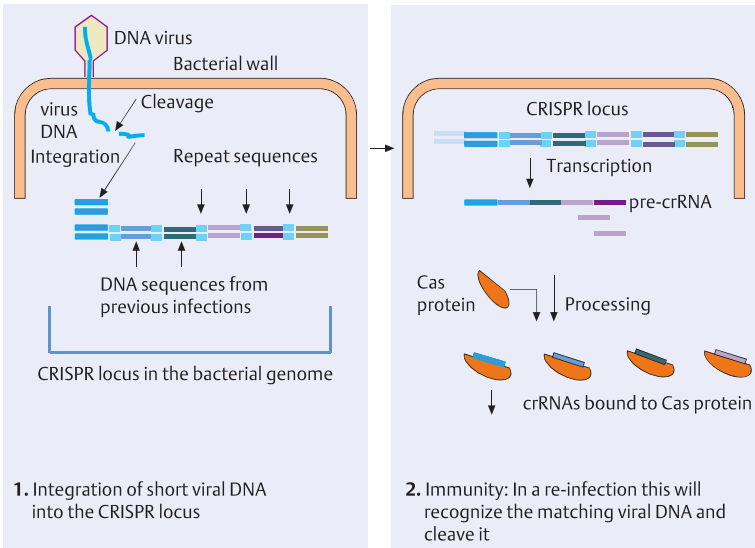
Mali P, et al. Cas9 as a versatile tool for engineering biology. Nat Methods 2013; 10(10): 957–963

Petherick A. Genome editing. Nature 2015; 528 (7580): S1

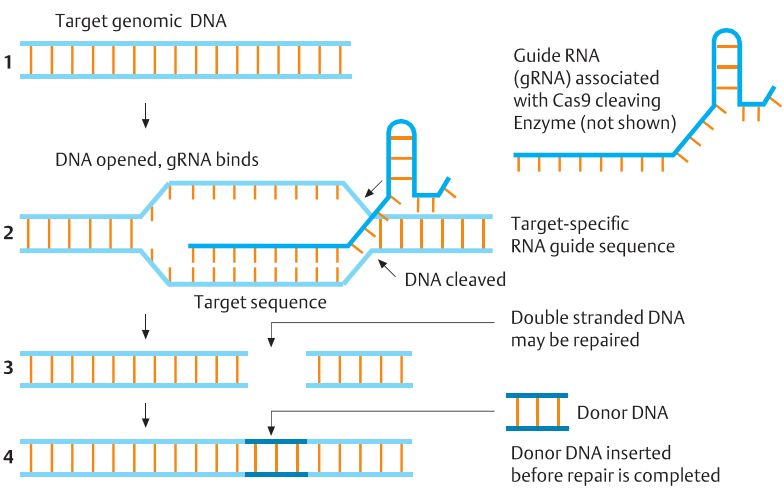
Sternberg SH, et al. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 2014; 507 (7490): 62–67

Strachan T, et al. Genetics and Genomics in Medicine. New York, NY: Garland Science, 2015

van der Oost J, et al. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol 2014; 12(7): 479–492



A. Function of a CRISPR-Cas system in bacteria



B. Use of the CRISPR/Cas-system in genome editing

Evolution of Genes and Genomes

Genes and genomes existing today are the cumulative result of evolutionary events that have taken place in the past. The classical theory of evolution, as formulated by Charles Darwin in 1859, states that (i) all living organisms today have descended from organisms living in the past; (ii) organisms that lived during earlier times differed from those living today; (iii) the changes occurred more or less gradual, with only a small change at a time; and (iv) the changes usually led to divergent organisms, with the number of ancestral types of organisms being smaller than the number of types today.

A. Gene evolution by duplication

Studies of the various genomes indicate that different types of duplication must have occurred: duplications of individual genes or parts of genes (exons), subgenomic duplications, and, rarely, duplications of the whole genome (Ohno, 1970). Duplication of a gene relieves the selective pressure on that gene. After a duplication event, the gene can accumulate mutations without compromising the original function, provided the duplicated gene has separate regulatory control. The duplication can also result in a genetic dead end. In this case, mutations inactivate the gene, leading to a pseudogene. Each human chromosome harbors almost as many pseudogenes as functional genes.

B. Gene evolution by exon shuffling

The exon/intron structure of eukaryotic genes provides great evolutionary versatility. New genes can be created by placing parts of existing genes into a new context, using functional properties in a new combination. This is referred to as exon or domain shuffling.

C. Evolution of chromosomes

Evolution also occurs by structural rearrangements of the genome at the chromosomal level. Related species, for example, mammals, differ in the number of their chromosomes and chromosomal morphology, but not in the number of genes, which are often conserved to a remarkable degree. The human chromosome 2 appears to have evolved from the fusion of two primate chromosomes. The differences in chromosome 3 are much more subtle. The chromosome 3 of orangutan differs from that of man and other primates by a pericentric inversion. The banding patterns of all primate chromosomes are remarkably similar. This reflects their close evolutionary relationship. (Figure adapted from Yunis & Prakash, 1982).

D. Molecular phylogenetics and evolutionary tree reconstruction

A phylogenetic tree can be reconstructed based on different types of evidence: fossils, differences in proteins, immunological data, DNA–DNA hybridization, and DNA sequence similarity. The number of events that have to have taken place to explain the diversity observed today is determined. In the path from an ancestral gene (**1**), two events are shown schematically. Two categories of homologs are distinguished: *paralogs* and *orthologs* (**2**). Paralogs are homologous genes that have evolved by duplication within a species. Orthologs are genes present in different organisms that have evolved by vertical descent from an ancestral gene between different species. The human α-globin and δ-globin loci are examples of paralogs. The α-globin and β-globin genes of humans and other mammals are examples of orthologs. The adjective *paralogous* refers to nucleotide sequence comparisons.

Further Reading

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Darwin C. The Origin of Species by Means of Natural Selection. London: John Murray, 1859

Eichler EE, Sankoff D. Structural dynamics of eukaryotic chromosome evolution. Science 2003; 301 (5634): 793–797

Gilbert W. The exon theory of genes. Cold Spring Harb Symp Quant Biol 1987; 52: 901–905

Miller W, et al. Comparative genomics. Annu Rev Genomics Hum Genet 2004; 5: 15–56

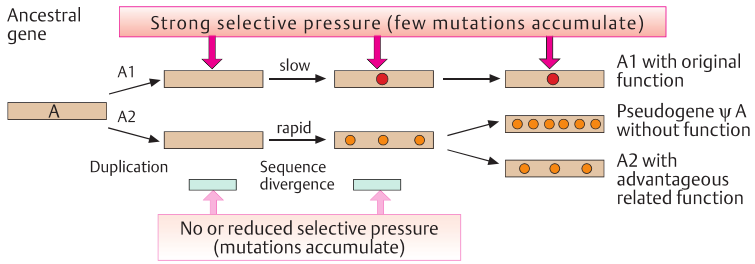
Jobling M, et al. Human Evolutionary Genetics. 2nd ed. New York, NY: Garland Science, 2014

Ohno S. Evolution by Gene Duplication. Heidelberg: Springer-Verlag, 1970

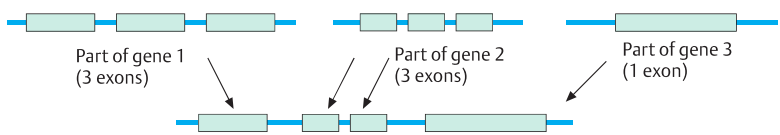
Rieseberg LH, Livingstone K. Chromosomal speciation in primates. Science 2003; 300(5617): 267–268

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Yunis JJ, Prakash O. The origin of man: a chromosomal pictorial legacy. Science 1982; 215(4539): 1525–1530

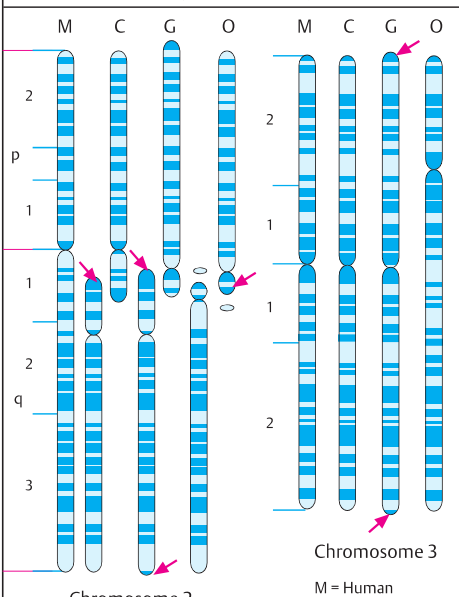


A. Gene evolution by duplication

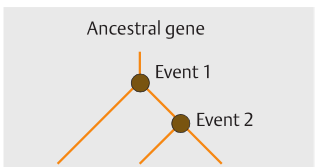


New gene with exons derived from three different genes and novel function

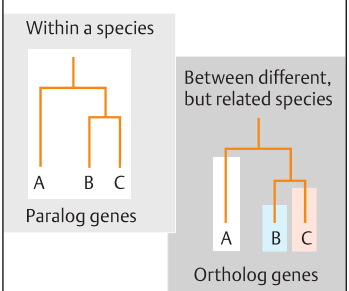
B. Gene evolution by exon shuffling



C. Evolution of chromosomes



1. Three genes related by evolution (gene tree)



2. Homologies of genes within and between species

D. Molecular phylogenetics and evolutionary tree reconstruction

Comparative Genomics

Comparative genomics refers to a systematic analysis of genomes of different species by comparing their similarities and differences. It compares DNA sequences, the structure and function of genes and their chromosomal location, regulatory sequences, and other features. Genomes of different organisms differ or resemble each other according to their evolutionary relationship. The number and types of differences depend on the time that has passed since their last common ancestor existed and the frequency of genomic changes and mutations that have occurred since they diverged from the ancestor. If two genomes are sufficiently closely related, the order of genes along each chromosome will have been preserved.

Two areas of comparative genomics have practical applications: plant genomes and the search for human disease genes that have homologs in other organisms.

A. Homologies of human proteins

The sequence of the human genome has revealed considerable homologies of its proteins with those of other organisms, for example, 21% with other eukaryotes and prokaryotes.

B. Chromosome-associated proteins

The chromatin-associated proteins and transcription factors are conserved in evolution. As a consequence, 60% of human, *Drosophila*, and the nematode *C. elegans* chromatin architectures are shared.

C. Chromosomal segments conserved between man and mouse

As a result of evolution from a common ancestor approximately 80 million years ago, the human genome contains approximately 340 segments that are shared with the mouse genome. Within these segments of average length of 10 Mb, with a range from 24 kilobases to 90.5 Mb (Lander et al., 2001), the order of homologous genes is similar. This is referred to as synteny (see p. 116). The figure shows segments containing at least two genes in conserved order between man and mouse, indicated in a color corresponding to the mouse chromosome.

Further Reading

Brown TA. Genomes. 3rd ed. New York, NY: Garland Science, 2007

Lander ES, et al; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409 (6822): 860–921

Margulies EH, Birney E. Approaches to comparative sequence analysis: towards a functional view of vertebrate genomes. Nat Rev Genet 2008; 9(4): 303–313

O'Bleness M, et al. Evolution of genetic and genomic features unique to the human lineage. Nat Rev Genet 2012; 13(12): 853–866

Rubin GM, et al. Comparative genomics of the eukaryotes. Science 2000; 287(5461): 2204–2215

Strachan T, Read A. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Touchman J. Comparative genomics. Nat Edu Knowl 2010; 3(10): 13–18

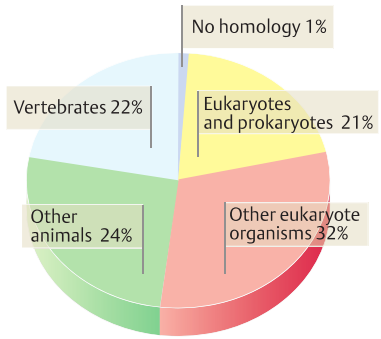
Waterston RH, et al. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature 2002; 420(6915): 520–562

Xia X. Comparative Genomics. Berlin: Springer, 2013

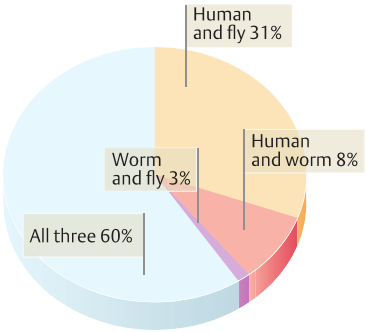
Comparative genomics of four nonvertebrate organisms based on genome sequence data

Organism	Number of genes	Genes arisen by duplication	Gene families (core proteome)	Type of organism
Haemophilus influenzae	1,709	284	1,425	Bacterium
Saccharomyces cerevisiae	6,241	1,858	4,383	Yeast
Caenorhabditis elegans	18,424	8,971	9,453	Nematode
Drosophila	13,601	5,536	8,065	Insect

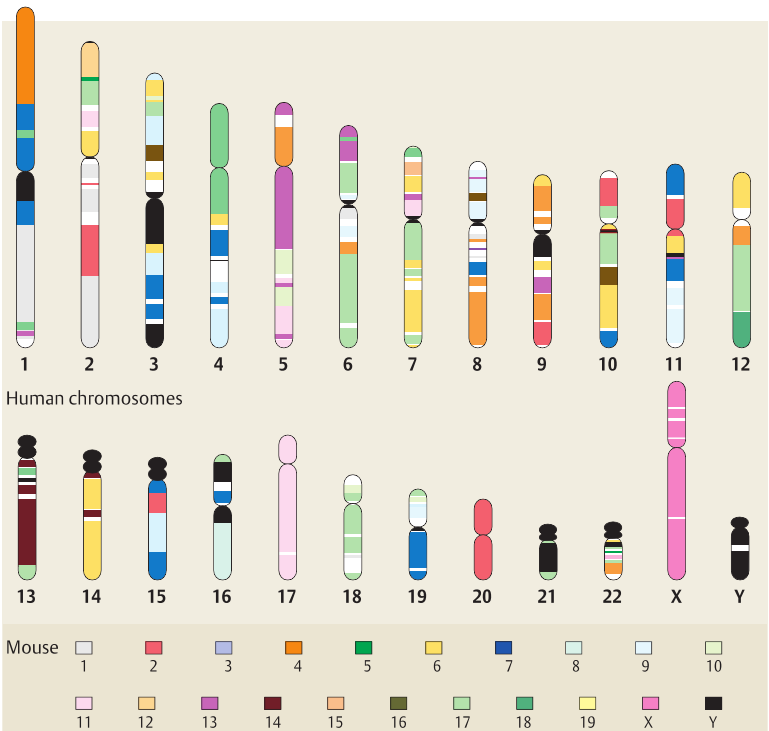
Data from Rubin et al. (2000).



A. Homologies of human proteins with those of other organisms



B. Homologies of chromosome-associated proteins



C. Chromosomal segments conserved between man and mouse

Genomic Structure of the Human X and Y Chromosomes

The human X and Y chromosomes have evolved from a pair of ancestral chromosomes during the past 300 million years. While the X chromosome retained many properties of an autosome, the Y chromosome has lost most of its genes and become greatly reduced in size. Its genetic function is now limited to inducing male development during embryonic development and to maintaining spermatogenesis in adult males. The two chromosomes undergo pairing and recombination at the distal ends of their short arms in the *pseudoautosomal region* (PAR1), whereas all other regions are exempted from recombination.

A. Genomic structure of the human X chromosome

Functional genes are distributed along the X chromosome with 153 million base pair (Mb) DNA, as shown by the dark blue squares (each square represents one gene). The approximate locations of nine selected landmark genes and their directions of transcription (arrows) are shown to the right. Most of the short arm consists of a region, the X-added region (XAR). This has resulted from the translocation of an ancestral autosome into the short arm (Xp) approximately 105 million years ago. The long arm (Xq) is composed of a region, XCR (X-conserved region), which is derived from an ancestral autosome. XCR has been conserved during evolution in mammals. Five regions of evolutionary conservation, evolutionary strata S1 to S5, have been identified along the X chromosome. The human X chromosome contains 1,098 genes with 7.1 genes per million base pairs, one of the lowest gene densities in the human genome (average: 10–13). (Figure adapted from Ross et al, 2005).

B. Genomic structure of the human Y chromosome

The human Y chromosome with 59-Mb DNA encodes approximately 72 protein-coding genes. It has a distinct genomic structure comprising five different regions in the euchromatic part: (i) two pseudoautosomal regions at the distal ends of the short (PAR1) and long arms (PAR2); (ii) the Y-specific male determinant (MSY) region of approximately 35 kilobases; (iii) approximately 8.6 Mb (38% of the euchromatic portion) called *X-degenerate*, derived from an ancestral autosome; (iv) 3.4 Mb derived from former X-linked genes by transposition (*X-transposed*), which occurred approximately 3 to 4 million years ago; and (v) 10.2 Mb amplified (*amplionic*) Y-specific sequences. The latter are designated P1 to P8 and are derived from former X- and Y-linked genes. They are termed *amplionic* because they consist of amplified palindromic sequences (amplicons) of various sizes with a marked sequence similarity of 99.9% over long stretches of DNA (tens to hundreds of kilobases). Most genes in the amplionic segments are expressed exclusively in testes.

Since the male-specific sequences on the Y chromosome do not participate in crossing-over, they are deprived of a mechanism for replacing mutations or structural rearrangements with normal sequences. Gene conversion between these palindromic sequences (Y–Y conversion) presumably serves as a mechanism for restoring normal sequences that have been rendered nonfunctional in one arm of a palindrome. (Figure adapted from Skaletsky et al, 2003).

C. Homologies between the X and Y chromosomes

The X and Y chromosomes share regions of homology as a result of their common evolutionary origin (for details see Ross et al., 2005).

Medical relevance

Four regions of Yq are associated with male infertility: AZFa, AZFb, AZFc, and DAZ (OMIM 415000, 400003, 400026/27).

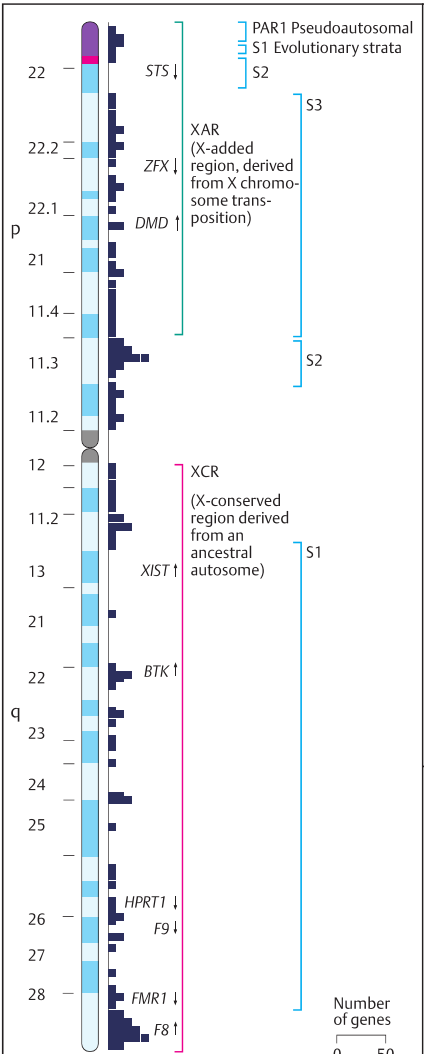
Further Reading

Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 2003; 4(8): 598–612

Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. Science 1999; 286(5441): 964–967

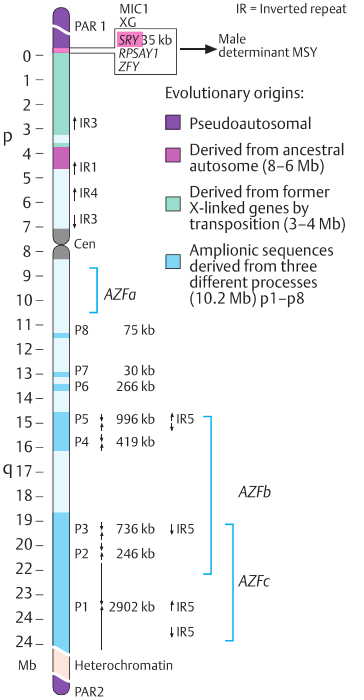
Ross MT, et al. The DNA sequence of the human X chromosome. Nature 2005; 434(7031): 325–337

Skaletsky H, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 2003; 423(6942): 825–837

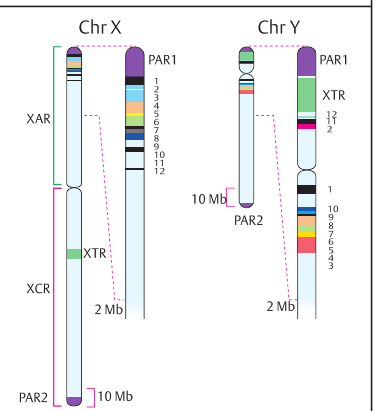


Selected landmark genes:
STS: steroid sulfatase, *ZFX*: Zinc finger protein, *XIST*: X-inactivation-specific transcript, *BTK*: Bruton agammaglobulinemia tyrosine kinase, *HPRT1*: Hypoxanthine-phosphoribosyltransferase 1, *F9* and *F8*: Hemophilia A and B, *FMR1*: Familial Mental Retardation type 1.

A. Genomic structure of the X chromosome



B. Genomic structure of the Y chromosome



C. Homologies of the X and Y chromosome

The Mitochondrial Genome of Man

The human genome contains an extranuclear genome derived from a transfer of a bacterial genome into a eukaryotic cell approximately 2×10^9 years before the present day, the mitochondrial genome. Each human cell contains hundreds of mitochondria. The mitochondrial genome in mammals is small and contains no introns, and in some regions, the genes overlap, and therefore practically every base pair is part of a coding gene. In germ cells, mitochondria are almost exclusively present in oocytes, whereas spermatozoa contain few. Thus, mitochondria are inherited exclusively from the mother (maternal inheritance), without paternal contribution.

A. Mitochondrial genes in man

The size of the human mitochondrial genome, sequenced in 1981 by Andersen and coworkers is only 16.5 kilobases (16,569 base pairs). Each mitochondrion contains 2 to 10 DNA molecules. In the two strands of the mitochondrial DNA (mtDNA), a heavy (H) strand and a light (L) single strand can be differentiated by a density gradient. Human mtDNA contains a total of 37 genes essential for intracellular energy transfer and conversion. Of these genes, 13 are protein-coding genes involved in four metabolic processes: (i) NADH (nicotinamide adenine dinucleotide) dehydrogenase, (ii) the cytochrome c oxidase complex (subunits 1–3), (iii) for cytochrome *b*, and (iv) for subunits 6 and 8 of the ATPase (adenosine triphosphatase) complex. Unlike that of yeast, mammalian mtDNA contains seven subunits for NADH dehydrogenase (ND1, ND2, ND3, ND4L, ND4, ND5, and ND6).

Most genes are found on the H strand. The L strand codes for a protein (ND subunit 6) and eight transfer RNAs (tRNAs). From the H strand, 14 tRNAs and two RNAs are transcribed: a short one for the 12S ribosomal RNA (rRNA) and a long one for 16S rRNA. A single transcript is made from the L strand. A 7S RNA is transcribed in a counterclockwise manner close to the origin of replication (ORI), located between the 11 and 12 o'clock on the circular structure.

B. Cooperation between mitochondrial and nuclear genome

Many mitochondrial proteins are aggregates of gene products of nuclear and mitochondrial genes. Approximately 1,100 nuclear genes cooperate with the mitochondrial genome. In the mitochondria, they form functional proteins from subunits of mitochondrial and nuclear gene products. This explains why several mitochondrial genetic disorders show Mendelian inheritance, whereas purely mitochondrially determined disorders show exclusively maternal inheritance.

C. Evolutionary relationship of mitochondrial genomes

Mitochondria probably evolved from independent prokaryotic organisms that were integrated into eukaryotic cells.

Medical relevance

Numerous maternally transmitted and de novo diseases result from structural changes and point mutations in transfer DNA. (see Part III, Mitochondrial Diseases, p. 260, and Appendix, Table 8, p. 409).

Further Reading

Anderson S, et al. Sequence and organization of the human mitochondrial genome. Nature 1981; 290 (5806): 457–465

Genetics Home Reference. Mitochondrial DNA. Available at http://ghr.nlm.nih.gov/chromosome/MT. Accessed July 25, 2017

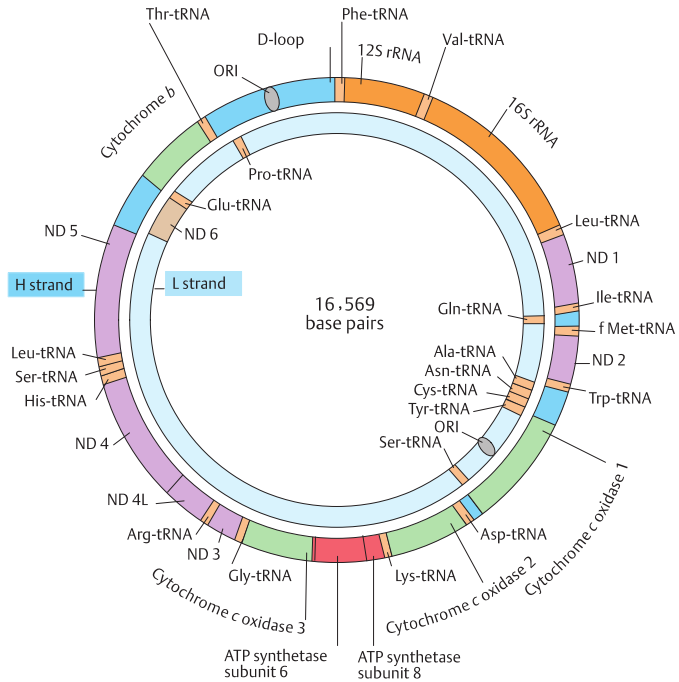
Schon EA, et al. Human mitochondrial DNA: roles of inherited and somatic mutations. Nat Rev Genet 2012; 13(12): 878–890

Wallace DC. Genetics: mitochondrial DNA in evolution and disease. Nature 2016; 535(7613): 498–500

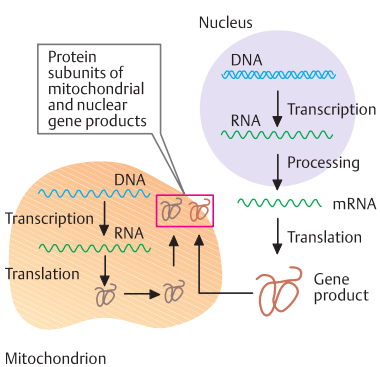
Wallace DC. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Annu Rev Genet 2005; 39: 359–407

Wallace DC, Fan W. Energetics, epigenetics, mitochondrial genetics. Mitochondrion 2010; 10(1): 12–31

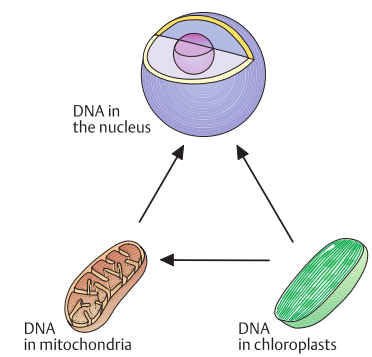
Wallace DC, et al. Mitochondrial medicine: the mitochondrial biology and genetics of metabolic and degenerative diseases, cancer, and aging. In: Rimoin DL, Pyeritz RE, Korf BR, eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013 (E-book only)



A. Mitochondrial genes in man



B. Cooperation between mitochondrial and nuclear genome



C. Evolutionary relationship of mitochondrial genomes

Genetics in Medicine

Genomic Disorders

Genomic disorders are a group of approximately 70 phenotypically distinct disorders resulting from rearrangements of the underlying structure of the human genome rather than mutations in genes. Copy number variations (CNVs), such as segmental duplications, play an important role (p. 210). Their sequence identity predisposes to nonhomologous pairing during meiosis and subsequent nonallelic recombination (NAHR). The resulting duplication/deficiency in the daughter cells causes a local genomic imbalance, which leads to developmental disturbances and disease.

A. Peripheral neuropathy

Peripheral hereditary motor-sensory neuropathy type 1A (HMSN1A; OMIM 118220; Charcot–Marie–Tooth disease, CMT1A) is the most frequent (1:2,500–7,500) neurogenic muscle atrophy. It begins in the first two decades of life and progresses with painful sensory perceptions, dissipation of muscle reflexes, and peripheral muscle atrophy of the feet, lower extremities, and hands. A reduced nerve conduction velocity of less than 38 m/s establishes the diagnosis.

B. Duplication/deletion as cause of different neuropathies

The underlying causes of two clinically distinct neurogenic disorders are two low-copy repeat segmental duplications in the distal short arm of chromosome 17 flanking a gene (**1**). The proximal repeat and the distal repeat are separated by 1.4 Mb genomic DNA at 17p11.2. A tandem duplication (**2**) involving the *PMP22* gene encoding peripheral myelin protein-22 (OMIM 601097) is the cause of CMT1A in approximately 50% of patients. This duplication occurs *de novo* in 75 to 90% of patients. Point mutations have also been observed. An allelic, reciprocal disorder is hereditary neuropathy with liability to pressure palsies (HNPP; OMIM 162500) due to deletion involving gene *PMP22* at 17p11.2 (**3**). The PMP22 protein is a component of myelinated peripheral nerve fibers. OMIM 118220 lists 70 different forms of Charcot–Marie–Tooth disease.

C. Nonallelic homologous recombination

Segmental duplications (SD1 and SD2) flanking the genomic region 17p11.2 (**1**) predispose to nonhomologous pairing (**2**) during meiosis because of their high sequence homology. Nonallelic homologous recombination (NAHR) of two segmental duplications results in two reciprocal states of imbalance: a duplication (**3**) in one daughter cell and a deletion in the other daughter cell (**4**). In the case of CMT1A, the majority of crossovers occur within a limited region known as a recombination hotspot. Numerous other genomic disorders have been recognized.

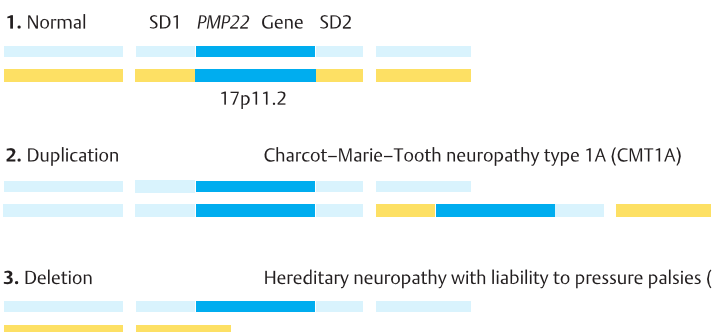
Further Reading

Boone PM, et al. Genomic medicine and neurological disease. Hum Genet 2011; 130(1): 103–121
Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 2016; 17(4): 224–238
Cooper GM, et al. A copy number variation morbidity map of developmental delay. Nat Genet 2011; 43 (9): 838–846
Girirajan S, et al. Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet 2010; 19(R2): R176–R187
Klopocki E, et al. Copy-number variations, noncoding sequences, and human phenotypes. Annu Rev Genomics Hum Genet 2011; 12: 53–72
Liu P, et al. Mechanism, prevalence, and more severe neuropathy phenotype of the Charcot-Marie-Tooth type 1A triplication. Am J Hum Genet 2014; 94(3): 462–469
Lupski JR, et al. Genomic Disorders. The Genomic Basis of Disease. Totowa, NJ: Humana Press, 2006
Lupski JR, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med 2010; 362(13): 1181–1191
Vissers LE, et al. Microdeletion and microduplication syndromes. Methods Mol Biol 2012; 838: 29–75

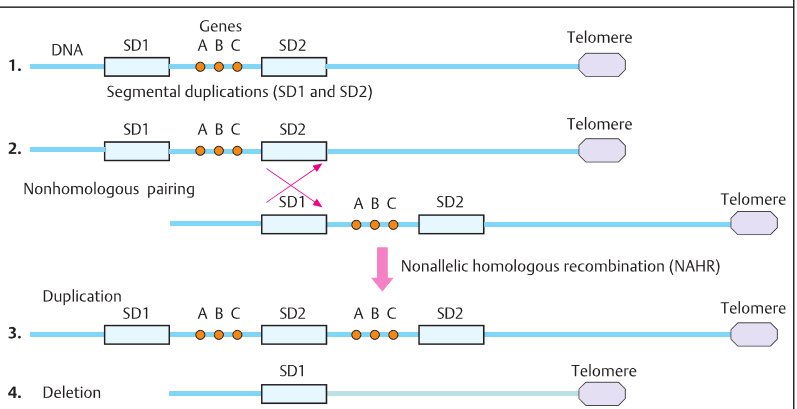
Examples of other genomic disorders

Smith–Magenis syndrome (SMS; OMIM 182290) 17p11.2 deletion
Duplication syndrome (17)(p11.2p11.2) 17p11.2 duplication
Deletions causing neurofibromatosis type I (NF1; OMIM 162200) 17q11.2 deletion
Sotos syndrome (OMIM 117550) 5q35 deletion
Many microdeletions and microduplications (p. 386)

A. Peripheral neuropathy type 1A



B. Duplication/deletion as cause of different neuropathies



C. Nonallelic homologous recombination

Disorders due to Dysregulated Chromatin Structure

More than approximately 20 genetically defined human disorders result from dysregulated chromatin structure, collectively called chromatin disorders or *chromatinopathies*. They result in developmental delay often associated with a recognizable phenotype and intellectual disability.

A. Two examples of chromatin disorders

Coffin–Siris syndrome (**1**, CSS, OMIM 135900), first described in 1970, is a multiple malformation syndrome of intellectual disability associated with coarse facial features, hypertrichosis, sparse scalp hair (**1**), and hypoplastic or absent fifth fingernails or toenails (**2**). Retarded growth and other malformations occur. At least five types of CSS are known, caused by mutations in five genes (see part C). The gene most frequently affected in 28 to 47% of patients is *ARID1B* (614556) at chromosomal location 6q25.3. *Nicolaides–Baraitser syndrome* (**3**, NCBRS, 601358), first described in 1993, consists of severe intellectual disability, short stature, dysmorphic facial features with sparse hair growth, unusual hand morphology with short metacarpals (**4**), and early-onset seizures due to mutations in the *SMARCA2* gene (600014). (Figures courtesy Professor Dagmar Wieczorek, Düsseldorf.)

B. Regulation of chromatin remodeling

Chromatin remodeling (see next page) is regulated by several enzymes: histone acetyl transferases (HATs), histone deacetylases (HDACs), histone methyl transferases (KMTs), histone demethylases (KDMs), and DNA methyltransferases (DNMTs). Euchromatin occurs in two reversible states, closed with transcription in genes repressed (**1**) or open (**2**), with genes in active transcription (see p. 181). *Silenced genes* are associated with DNA methylation (5mC), H3 lysine trimethylation at position 9 (H3K9me3), or trimethylation of H3 lysine 27 (H3K27me3). *Actively transcribed genes* in open euchromatin are associated with histone acetylation (H3/H4Kac) and trimethylation of H3 lysine 4 (H3K4me3) at promotors, and trimethylation of H3 lysine 36 (H3K36me3). The patterns of histone modification and DNA methylation are distinguished as “writers” (adding chemical groups), “erasers” (removing them), and “readers” (DNA- or histone-binding molecules that interpret epigenetic marks). (Figure modified from Mirabella et al., 2016.)

C. Mammalian SWI/SNF complex

A regulatory protein complex in chromatin, named SWI/SNF, consists of multiple subunits in subtypes BAF and PBAF involved in chromatin remodeling (see next page). It can alter the position of nucleosomes (see p. 138). Its name is derived from Switch/Sucrose Nonfermentable, a yeast and Drosophila homolog. Several *SMARC* genes (SWI/SNF-related, matrix-associated, actin-dependent regulators of chromatin) are involved in the etiology of CSS or NCBRS mentioned in part A. BAF subunits are frequently mutated in cancer. (Figure modified from Helming et al., 2014.)

Further Reading

Criscione SW, et al. The chromatin landscape of cellular senescence. Trends Genet 2016; 32(11): 751–761

Helming KC, et al. Vulnerabilities of mutant SWI/SNF complexes in cancer. Cancer Cell 2014; 26(3): 309–317

Mirabella AC, et al. Chromatin deregulation in disease. Chromosoma 2016; 125(1): 75–93

Kadoch C, et al. Mammalian SWI/SNF chromatin remodeling complexes and cancer: mechanistic insights gained from human genomics. Sci Adv 2015; 1(5): e1500447

Kosho T, et al. Coffin-Siris syndrome and related disorders involving components of the BAF (mSWI/SNF) complex: historical review and recent advances using next generation sequencing. Am J Med Genet C Semin Med Genet 2014; 166C(3): 241–251

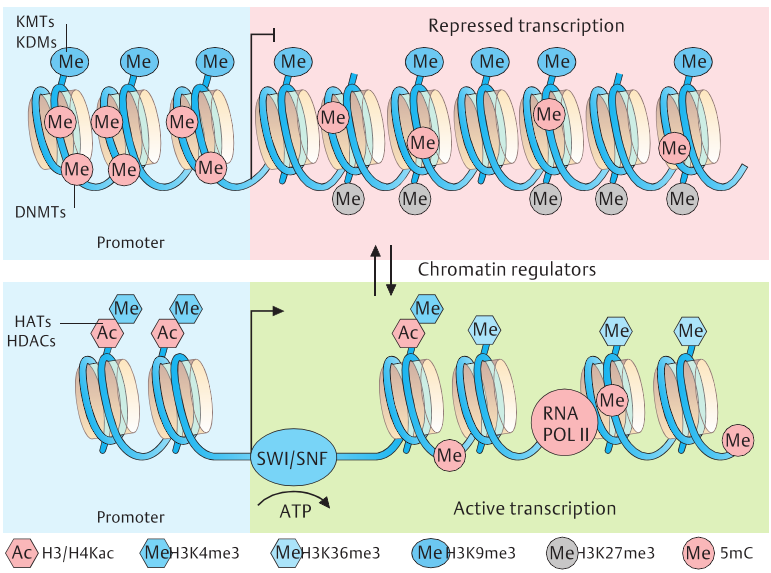
Sousa SB, et al. Nicolaides-Baraitser Syndrome International Consortium. Phenotype and genotype in Nicolaides-Baraitser syndrome. Am J Med Genet C Semin Med Genet 2014; 166C(3): 302–314

Tang L, et al. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. Prog Biophys Mol Biol 2010; 102(2–3): 122–128

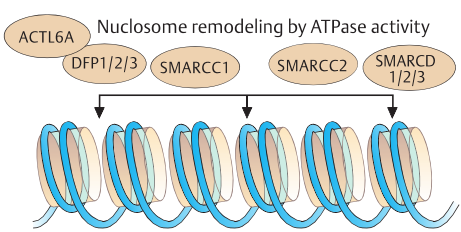
Wieczorek D, et al. A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. Hum Mol Genet 2013; 22(25): 5121–5135

Coffin-Siris syndrome (*SMARCB1* mutation) Nicolaides-Baraitser syndrome

A. Two examples of disorders of chromatin structure



B. Regulation of chromatin remodeling



C. Mammalian SWI/SNF complex

Genes involved in Coffin-Siris syndrome: *ARDID1A* (614607), *ARID1B* (614556), *SMARCB1*(601607), *SMARCA4* (603254), *SMARCE1* (603111).

In Nicolaides-Baraitser syndrome: *SMARCA2* (600014). Mutations in the others occur in different forms of cancer

Disorders Resulting from Rearrangement of *Cis*-Regulatory Elements

Structural aberrations involving *cis*-regulatory elements can result in human disorders. Main phenotypic manifestations are tissue- and cell-specific inborn errors of development leading to organ-specific congenital malformations. More than one gene in two adjacent topological-associated domains (TADs; see next page) may be involved. Deletions, duplications, or inversions within a TAD may interfere with normal enhancer–promoter interaction with loss of an insulator element, leading to wrong expression of a target gene.

A. Skeletal malformation due to a rearranged TAD

This example illustrates a rearrangement in a TAD (**1**) involving the gene *BMP2* located within a TAD at 20p12.3. It encodes the bone morphogenetic protein 2 (112261). This TAD is separated from neighboring TADs by boundary elements that function as insulators on both sides. The *BMP2* gene in the middle of the diagram (**2**) is normally transcribed under the influence of many enhancer elements located within the BMP2 TAD. One of these enhancers (a *cis*-acting regulatory element) has been located approximately 110 kb downstream in a 5.5-kb region (shown as red dot). Normally, the *BMP2* gene is transcribed (**3**, red arrow) under the control of a transcription factor when the enhancer (red dot) comes in contact with the gene in its promoter region. Normal *BMP2* function together with other genes leads to normal development of a hand as seen in the X-ray (**4**). In contrast, a duplication of the enhancer (**5**) is associated with an increased rate of transcription (thick red arrow). This causes a shortened index finger (**6,** white arrow) as a manifestation of brachydactyly type A2 (112600).

BMP2 is member of the large transforming growth factor beta (TGFβ) superfamily (see TGFβ signal pathway, p. 190) with a role in limb development. Together with SOX9 (608160) and WNT1 (164820), it modulates a network controlling digit development. (Figure adapted from Spielmann and Mundlos, 2013.)

B. Different types of structural rearrangements

Under normal circumstances (**1**) different TADs are separated by an insulator element serving as a functional boundary. In the schematic example shown, gene 2 in TADb is not transcribed (red bar), whereas in TADa gene 1 is transcribed under the influence of an enhancer on its promoter (green arrow). By an *inversion* of DNA sequences in this region (**2**), an enhancer (E) may be moved out of TADa (on the left, red triangle) and brought into the vicinity of gene 2 (G2) with the boundary (B) now on its right side. As a result, gene 2 is transcribed (red arrow) instead of gene 1 (green bar). In a *duplication* (**3**), the enhancer placed next to the duplicated region (G2′—B′—E′) causes transcription of both gene 2 (red arrow) and gene 1 (green arrow). If the boundary (B) is removed by a *deletion* (**4**), the enhancer can interact with both genes. In a smaller deletion not involving the boundary B (**5**, Deletion^{S}), the TAD structure remains unchanged and the enhancer can interact with the promoter of gene 1. Enhancer variants in a gene regulatory network cause a complex disorder, Hirschsprung disease (see p. 250; Neurocristopathies). (Figure adapted from Lupiáñez et al., 2015.)

Further Reading

Erickson RP, et al., eds. Epstein's Inborn Errors of Development. 3rd ed. Oxford: Oxford University Press, 2016

Franke M, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature 2016; 538(7624): 265–269

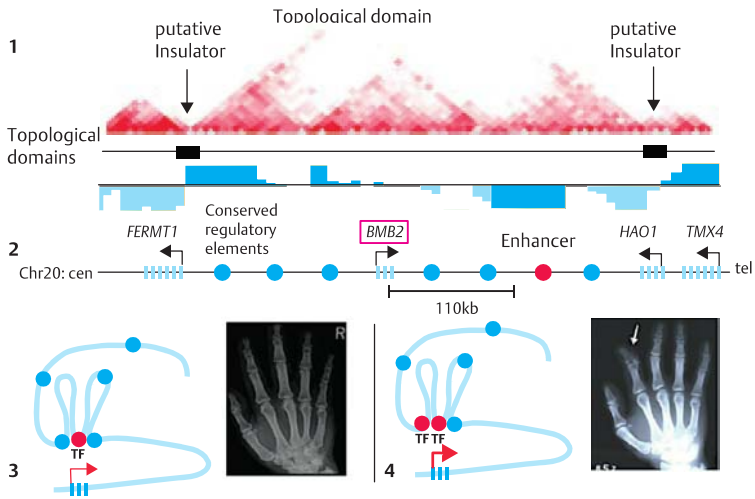
Mundlos S. The brachydactylies: a molecular disease family. Clin Genet 2009; 76(2): 123–136

Lupiáñez DG, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 2015; 161: 1012–1015.

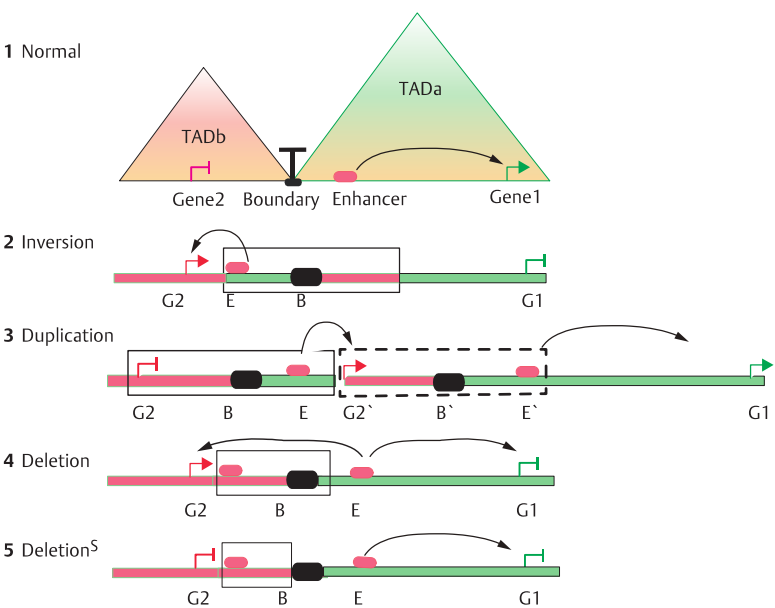
Spielmann M, Mundlos S. Structural variations, the regulatory landscape of the genome and their alteration in human disease. BioEssays 2013; 35(6): 533–543

Spielmann M, et al. Looking beyond the genes: the role of non-coding variants in human disease. Hum Mol Genet 2016; 25(R2): R157–R165

Stankiewicz P, et al. Structural variation in the human genome and its role in disease. Annu Rev Med 2010; 61: 437–455



A. Example of a skeletal malformation due a rearranged TAD



B. Different types of structural rearrangements

Disorders Resulting from Defects in Telomeres

The progressive loss of DNA at each cell division leads to telomere shortening, which eventually interferes with telomere stability. Telomere maintenance requires telomerase reverse transcriptase (TERT), a telomerase RNA component (TERC), and several other proteins. The shelterin proteins prevent telomeres from being processed as DNA double-strand breaks. Mutations in the genes encoding these proteins or the RNA component result in impaired telomere maintenance and cause a group of disorders collectively called telomere disorders or *telomeropathies*.

A. Dyskeratosis congenita, a telomere disorder

Dyskeratosis congenita (DKCA1; OMIM 127550), first described in 1910, is a genetically heterogeneous multisystem disorder resulting from mutations in seven genes encoding components of telomerase RNA. Its variable clinical features include abnormal skin pigmentation (**1**), dystrophy of the finger nails (**2**), white spots on the tongue (**3**) and inside the cheeks, bone marrow failure, pulmonary fibrosis, signs of early aging, and other manifestations. Several genetic types (DKC 1–7) are due to mutations in different genes. DKCA1 is caused by a heterozygous mutation in TERC (602322) on chromosome 3q26.2 in approximately 50% of DKCA1 patients. Mutations, either autosomal dominantly or recessively inherited, in several other genes that encode the proteins in part B are located at 5p15.32, 14q12, 16p22.1, 17p13.1, and 20q13.33. An X-linked form of DKC located at Xq28 is caused by mutation in the gene for dyskerin (300126). Hoyeraal–Hreidarsson syndrome (HHS, 305000) is an especially severe form of DKC caused by several defective DKC-associated genes, including X-linked DAC1 (300126). (Photographs: Kirwan and Dokal, 2009; open access under CC BY 3.0 license.)

B. Components of telomere maintenance

A complex of six proteins called *shelterins* coat telomeres: telomere repeat-binding factor 1 and 2 (TERF1, TERF2), TERF1-interacting nuclear protein 2 (TIN2), TIN2-interacting protein (TPP1), protection of telomeres (POT1), and repressor/activator protein1 (RAP1). They regulate telomere elongation and prevent DNA damage response. Telomerase contains four proteins (dyskerin, NOP10, NHP2, and GAR1), the RNA template TERC, and reverse transcriptase TERT. The CST complex with three members, conserved telomere protection component 1 (CTC1), suppressor of cdc 13 (STN1), and telomeric pathways with STN1 (TEN1), assumes different roles not described here. The protein TCAB1 guides telomerase to the telomere ends. The proteins shown in color and bold face are implicated in telomere disorders. (Image: Armanios and Blackburn, 2012.)

C. Telomere length and disease

Manifestations of telomere syndromes are age-dependent. For example, four different telomere disorders (HHS, DKCA1, aplastic anemia [609135], and pulmonary fibrosis [614743]) are manifest at different ages. In affected families, anticipation may be observed (increasing severity in generations following each other). Telomerase activity is upregulated in most human cancers.

Further Reading

Armanios M. Syndromes of telomere shortening. Annu Rev Genomics Hum Genet 2009; 10: 45–61

Armanios M, Blackburn EH. The telomere syndromes. Nat Rev Genet 2012; 13(10): 693–704

Ballew BJ, et al. A recessive founder mutation in regulator of telomere elongation helicase 1, RTEL1, underlies severe immunodeficiency and features of Hoyeraal Hreidarsson syndrome. PLoS Genet 2013; 9(8): e1003695

Bertuch AA. The molecular genetics of the telomere biology disorders. RNA Biol 2016; 13(8): 696–706

Bessler M, et al. Dysfunctional telomeres and dyskeratosis congenita. Haematologica 2007; 92(8): 1009–1012

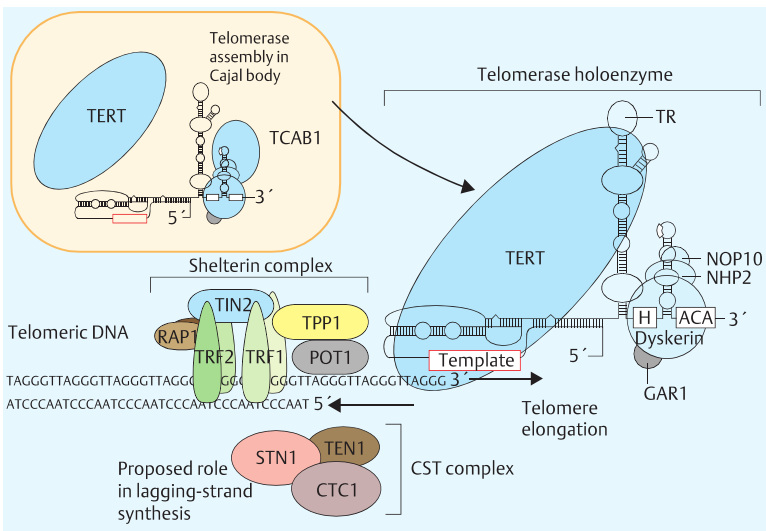
Holohan B, et al. Cell biology of disease: telomeropathies: an emerging spectrum disorder. J Cell Biol 2014; 205(3): 289–299

Kirwan M, et al. Dyskeratosis congenita, stem cells and telomeres. Biochim Biophys Acta 2009; 1792 (4): 371–379

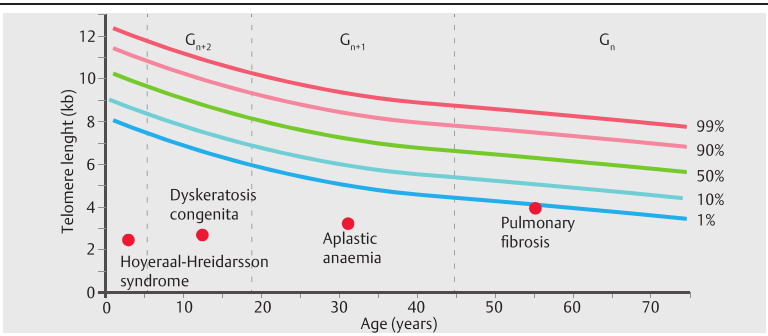
Savage SA, et al. The genetics and clinical manifestations of telomere biology disorders. Genet Med 2010; 12(12): 753–764

Townsley DM, et al. Bone marrow failure and the telomeropathies. Blood 2014; 124(18): 2775–2783

A. Dyskeratosis congenita, a telomere disorder



B. Components of telomere maintenance



C. Telomere length and disease

Disorders Resulting from Defective Lamins

Defective lamins cause of approximately 20 phenotypically and genetically different rare genetic disorders, collectively named *laminopathies*. Their phenotype can involve different tissues alone or in combination: muscle, bone, skin, nervous system, premature ageing disorders, and others (Table 5 in the Appendix, p. 407).

Lamins are proteins with high molecular weight (>400 kDA) of the basal lamina in the extracellular matrix. Lamins are heterotrimeric proteins (three chains: α, β, and γ) of approximately 15 types. Alpha-chain forms are named LAMA1, LAMA2, LAMA3, etc., beta-chain forms LAMB1, LAMB2, etc., and gamma-chain forms LAMC1, LAMC2, etc. Lamins are part of tissue scaffolding by interacting with a variety of other molecules required for maintaining tissue structure, such as the cell and nuclear membrane, and with functions in control of transcription and cell survival.

A. Three examples of laminopathies

Progeria, restrictive dermopathy, and Emery–Dreifuss muscular dystrophy are three examples which illustrate the strikingly different phenotypes of laminopathies. *Progeria* (**1**, OMIM 176670), also called Hutchinson–Gilford syndrome, is characterized by multiple signs of premature aging beginning within the first year of life. It is caused by a de novo heterozygous mutation in the lamin A gene *LMNA* (150330, see later). *Restrictive dermopathy* (**2**, OMIM 275210) is a lethal skin disease with thin translucent and tight skin, congenital contractures, intrauterine growth retardation, and other manifestations. Two different genetic forms occur due to mutations in the *LMNA* gene or in the *ZMPSTE24* gene (606480). *Emery–Dreifuss muscular dystrophy* (**3**, OMIM 310300) is a genetically heterogeneous adult-onset slowly progressive muscular dystrophy of the proximal upper and distal lower extremities (**4**), and the heart (**5**, life-threatening conduction defects). Several other genetic forms exist (see OMIM 310300 phenotypic series). (Figure in 1: Scaffidi et al., 2005; in 2: Sharma and Mahajan, 2011; in 3: Kornberg (www.neuromuscular.wustl.edu) and Stucki et al., 2000.)

B. Nuclear Lamina

The nuclear lamina (lamins A, B, and C) is located at the inner nuclear membrane. It binds to nuclear pore complexes, several types of lamina-associated proteins, such as emerin, and numerous other nuclear envelope proteins (not shown in this simplified diagram). The nuclear lamina maintains nuclear stability and relates to chromatin; thus, it is involved in regulatory functions. (Figure redrawn from Capell and Collins, 2006.)

C. Lamin A/C

The *LMNA* gene (OMIM 150330) located on chromosome 1q22 consists of 12 exons spread along 57.6-kb DNA. It encodes lamin A and lamin C (OMIM 150330) in approximately equal amounts. An alternative splice site in intron 10 produces lamin C, transcribed from exons 1 to 9 and part of exon 10. Lamins have three domains, an amino (N)-terminal globular domain, a central alpha-helical coiled-coil rod domain, and a globular carboxy (C)-terminal tail domain. Lamin homodimers associate with protofilaments, which form the 10-nm intermediate filament structure. Mutations can occur throughout the entire gene and cause one of the laminopathies (the site for only five examples are shown). (Figure redrawn from Capell and Collins, 2006.)

Further reading

Andrés V, et al. Role of A-type lamins in signaling, transcription, and chromatin organization. J Cell Biol 2009;187(7):945–957

Butin-Israeli V, et al. Nuclear lamin functions and disease. Trends Genet 2012;28(9):464–471

Capell BC, Collins FS. Human laminopathies: nuclei gone genetically awry. Nat Rev Genet 2006;7(12):940–952

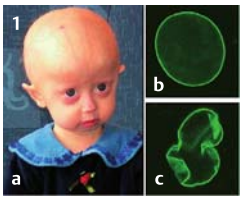
Navarro CL, et al. Loss of ZMPSTE24 (FACE-1) causes autosomal recessive restrictive dermopathy and accumulation of Lamin A precursors. Hum Mol Genet 2005;14(11):1503–1513

Scaffidi P, et al. The cell nucleus and aging: tantalizing clues and hopeful promises. PLoS Biol 2005;3(11):e395

Schreiber KH, et al. When lamins go bad: nuclear structure and disease. Cell 2013;152(6):1365–1375

Sharma S, et al. Collodion baby. Indian Dermatol Online J 2011;2(2):133

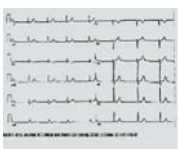
Stucki C, et al. Cardiopathy in a patient with Emery-Dreifuss muscular dystrophy. J Clin Basic Cardiol 2000;3(2):145–146



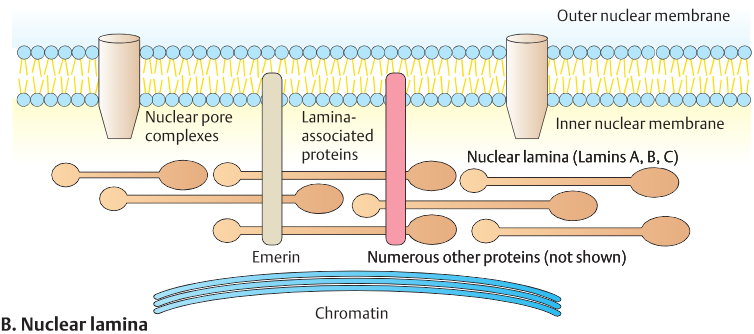
Progeria.
a) phenotype,
b) nuclear
Membrane
normal,
c) defective

Restrictive
dermopathy

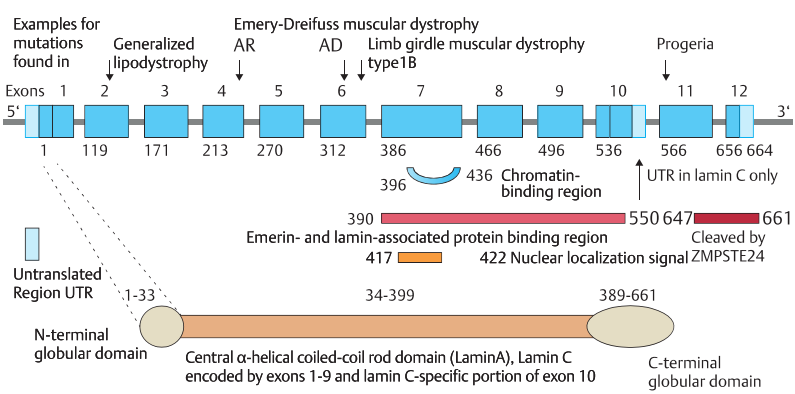
Emery-
Dreifuss
muscular
dystrophy



A. Three examples of laminopathy



B. Nuclear lamina



C. Lamin A/C

Disorders Resulting from Dysfunctional Cohesin

Two structurally and functionally related protein complexes, condensin and cohesion, maintain the structure of chromosomes during mitosis and meiosis (p. 92).

Cohesin holds the sister chromatids together after DNA replication until anaphase (see cell cycle, p. 90). Mutations in genes encoding one of the cohesion subunits cause a genetically heterogeneous group of disorders, named cohesinopathies. The prototype of these disorders is the Cornelia de Lange syndrome (CdLS).

A. Facial phenotype of Cornelia de Lange syndrome

CdLS (122470) is a developmental retardation syndrome with a wide spectrum of manifestations from severe to very mild. It can usually be recognized by its characteristic facial phenotype: arched eyebrows joined over the nasal bridge (synophrys), anteverted nostrils, long philtrum, and downturned corners of the mouth (**1–3**). Microcephaly with variable degree of retarded physical and mental development is common. In approximately 10% of patients, characteristic severe malformations of the upper extremities are present (**4**). This was described by Brachmann in 1916 as a different syndrome, but is now considered part of the spectrum of CdLS. Aside from rare exceptions CdLS occurs isolated without increased familial incidence.

B. Cohesin complex

Cohesin proteins form a cohesion complex consisting of four subunits. The cohesion proteins are distributed along each sister chromatid of a metaphase chromosome (**1**). Seven genes (**2**) encode the cohesion complex (**3**). Two cohesin subunits, SMC1A and SMC3 (structural maintenance of chromosomes), form a ring structure around DNA. Several other proteins contribute to the complex with additional functions including transcription regulation, DNA double-strand repair, chromosome condensation, homologous pairing, and others. (The spatial relationships of the proteins in the lower half of the diagram may differ from the positions shown.)

Mutations in five genes have been recognized as a cause of CdLS: *NIPBL*, *SMC1A*, *SMC3*, *RAD21*, and *HDAC8*. Approximately 50 to 60% of patients with CdLS are due to a mutation in the *NIPBL* gene (Nipped-B Drosophila homolog of de Lange, OMIM 608667). It encodes the delangin protein involved in cohesin loading onto DNA. Mutations in other components have been identified in 5 to 7% of the patients (5% in *SMC1A*, few in *SMC3* or *RAD21*). In 5% of the patients, mutations in *HDAC8* encoding an acetyltransferase of SMC3 are present. In 20% of the patients, no mutation is found. The types of mutations are nonsense, deletion/insertion, missense, and splice site. A severe phenotype is associated with nonsense mutations, whereas missense mutations are associated with a milder phenotype.

Related cohesinopathies are caused by mutations in *ESCO2* encoding a deacetylase of SMC3 (Roberts syndrome/SC phocomelia syndrome, 268300/269000). Warsaw breakage syndrome (613398) due to mutations in the *DDX11* gene (601150). (Professors Gabriele Gillessen-Kaesbach and Frank Kaiser, Lübeck, kindly provided advice, photographs, and the schematic figure in part B3.)

Further Reading

Boyle MI, et al. Cornelia de Lange syndrome. Clin Genet 2015; 88(1): 1–12

Deardorff MA, et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature 2012; 489(7415): 313–317

Deardorff MA, et al. Cohesinopathies. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier

de Lange C. Sur un type nouveau de degenerescence (typus Amstelodamensis). Arch Méd Enfants 1933; 36: 713–719

Mehta GD, et al. Cohesin: functions beyond sister chromatid cohesion. FEBS Lett 2013; 587(15): 2299–2312

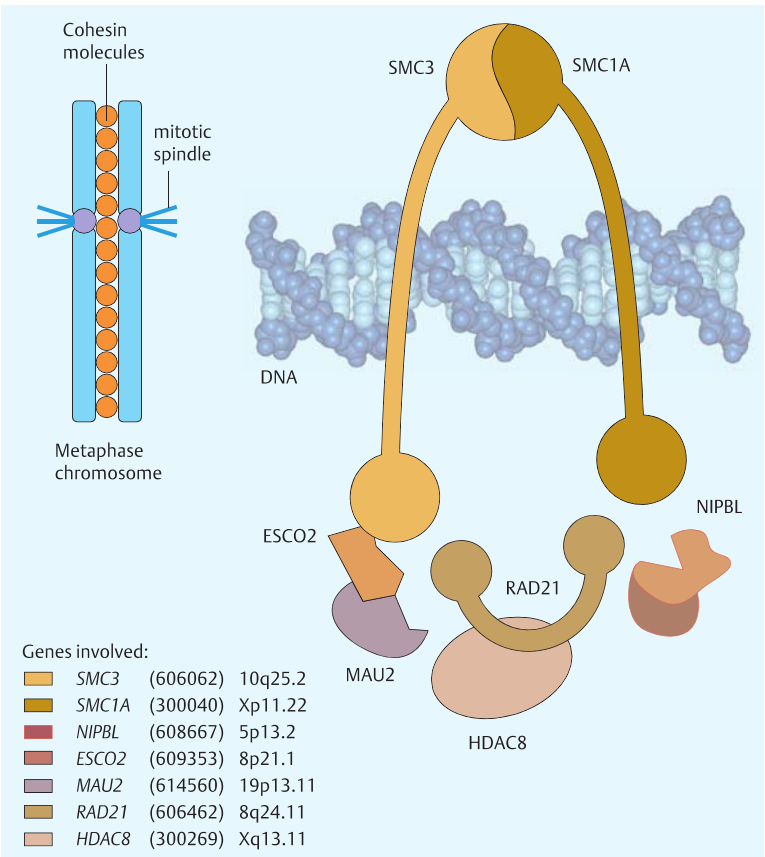
Krantz ID, et al. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet 2004; 36(6): 631–635

Liu J, et al. Cohesin and human disease. Annu Rev Genomics Hum Genet 2008; 9: 303–320

Mannini L, et al. Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome. Hum Mutat 2013; 34(12): 1589–1596

Wood AJ, et al. Condensin and cohesin complexity: the expanding repertoire of functions. Nat Rev Genet 2010; 11(6): 391–404

A. Facial phenotype of Cornelia de Lange syndrome



B. The cohesin complex

Disorders due to Dysfunctional Cilia (Ciliopathies)

Ciliopathies are a group of phenotypically different genetic disorders with a common pathogenic background. They share impaired function of cilia, a specific cell organelle. The wide phenotypic spectrum of ciliopathies results from the wide variety of structures and cells that harbor cilia: the respiratory tract, oviduct and spermatozoa, inner ear, photoreceptors, gallbladder, pancreas, kidney, ependymal lining of the brain, chondrocytes, and others. Cilia are highly conserved in evolution and are almost identical from unicellular green algae to primates.

A. Motile cilia: basic structure

All cilia consist of two parts: a basal body below the cell surface and an axoneme extending from the cell. Perpendicular to a central pair of tubuli that run the length of the cilium, nine radial spokes connect to nine peripheral microtubular doublets. Each doublet is a heterodimer of α and β tubulins assembled into A and B tubules, respectively. Microtubulin-associated proteins form an outer and an inner dynein arm (**1**). Two major classes of dyneins are known: axonemal and cytoplasmic. Axonemal dynein arms are composed of microtubule-associated motor protein complexes consisting of several heavy, light, and intermediate chains. The dynein arms are assembled in the cytoplasm before they are transported to the growing cilia (**2**). The dynein heavy chains derive energy for cilia motility from ATPase activity. Approximately 250 proteins are known to be required to make up a cilium. Intraflagellar transport (IFT) proteins serve to assemble and maintain the structure of cilia, but also are involved in signaling. (Figure from Omran and Olbrich, 2010, kindly provided by the authors.)

B. Primary ciliary dyskinesia

Primary ciliary dyskinesia (PCD), also known as Kartagener syndrome when associated with situs inversus, is a heterogeneous group of hereditary disorders resulting from mostly autosomal recessive mutations in more than 15 known genes (OMIM). Usually one or both of the dynein arm types are structurally defective. A typical defect occurs in the axonemal dynein arm. The outer dynein arm (ODA) can be seen in an electron microscopy cross-section of cilia in normal cells (**1**), whereas it is absent in mutant cells (**2**). Immunofluorescence microscopy of normal respiratory epithelial cells reveals an overlap of acetylated **α**-tubulin (green) and ODA (red) staining, resulting in a yellow coloration of the entire cilium in the normal cell (**3**), whereas ODA is lacking in the mutant cell (**4**). (Figure from Omran and Olbrich, 2010, kindly provided by the authors.)

C. Structure of the *DNAH5* gene

The *DNAH5* gene (OMIM 603335) on 5p14-p15 consists of 80 exons. Approximately 50% of individuals with PCD harbor a mutation in this gene.

Other examples of ciliopathies, including Bardet-Biedl syndrome, are listed in table 6 in the Appendix (p. 408).

Further Reading

Badano JL, et al. The ciliopathies: an emerging class of human genetic disorders. Annu Rev Genomics Hum Genet 2006; 7: 125–148

Cildb: a knowledgebase for centrosomes and cilia. Available at: http://cildb.cgm.cnrs-gif.fr/. Accessed February 27, 2017

Fliegauf M, et al. When cilia go bad: cilia defects and ciliopathies. Nat Rev Mol Cell Biol 2007; 8(11): 880–893

Horani A, et al. Genetics and biology of primary ciliary dyskinesia. Paediatr Respir Rev 2016; 18: 18–24

Jenkins D, et al. Genes and mechanisms in human ciliopathies. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

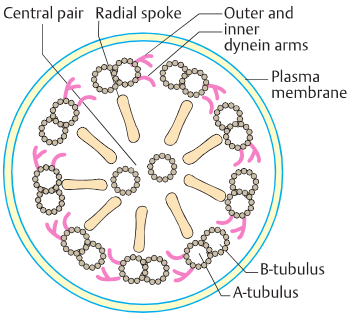
Madhivanan K, et al. Ciliopathies: the trafficking connection. Traffic 2014; 15(10): 1031–1056

Novarino G, et al. Modeling human disease in humans: the ciliopathies. Cell 2011; 147(1): 70–79

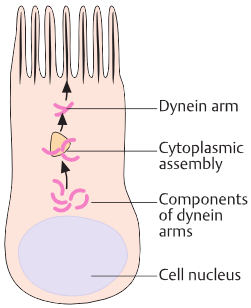
Omran H, Olbrich H. Zilienkrankheiten unter besonderer Berücksichtigung der primären ziliären Dyskinesie. Med Genetik 2010; 22(3): 315–321

Oud MM, et al. Ciliopathies: Genetics in pediatric medicine. J Pediatr Genet 2017; 6(1): 18–29

Witt M, et al. Primary ciliary dyskinesia (Kartagener syndrome). In: Erickson RP, et al., eds. Epstein's Inborn Errors of Development. 3rd ed. Oxford: Oxford University Press, 2016

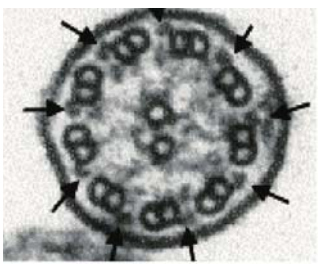


1. Motile respiratory cilia, cross-section

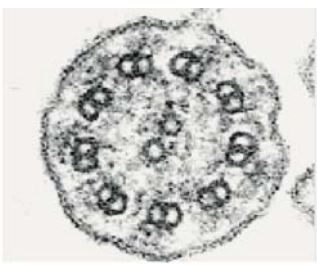


2. Respiratory epithelial cell

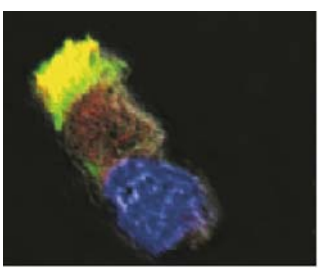
A. Motile cilia: basic structure



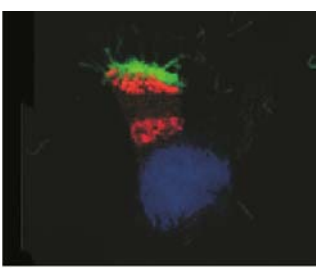
1.



2.

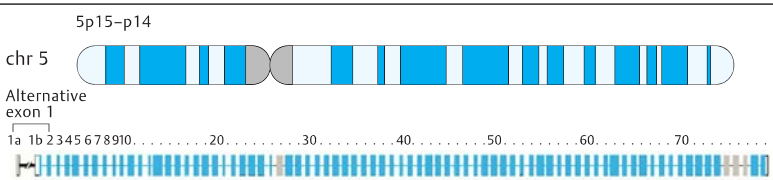


3.



4.

B. Primary ciliary dyskinesia (PCD)



C. Exon/intron structure of the *DNAH5* gene

Neurocristopathies

From the neural crest, first described by Wilhelm His in 1868 as *Zwischenstrang* (cord in between), diverse cell lineages of migratory cells develop into melanocytes, enteric neurons, and other cells. Dysregulated neural crest development resulting from mutations in different genes causes a group of different human diseases, the *neurocristopathies* (two examples in the following).

A. Multiple endocrine neoplasias

MEN2A (171400) and MEN2B (162300) are autosomal dominant disorders of endocrine tumors associated with other manifestations. The main tumor types are medullary thyroid carcinoma (MTC; shown in the figure), ganglioneuroma, and pheochromocytoma. Mutations in the *RET* proto-oncogene (164761) are the cause of the tumors, which arise de novo in approximately 50% of cases. MEN2A is an allelic disorder caused by missense mutations in *RET*. (Figures kindly provided by Professor K. W. Schmid, Essen.)

B. Hirschsprung disease (congenital intestinal aganglionosis)

Hirschsprung disease (HSCR; 142623), first described by H. Hirschsprung in 1888, is a multigenic neonatal intestinal obstruction syndrome correctable by surgery. It is genetically heterogeneous and clinically variable, ranging from severe in neonates (left photograph) to mild in late infancy and in childhood (middle photograph). The main clinical manifestations are chronic constipation, abdominal distension, and megacolon (right photograph) of variable length resulting from lack of gastrointestinal ganglion cells. The population incidence is 1 in 5,000 newborns and the sex ratio is approximately four males to one female. HSCR is classified according to the length of the aganglionic segment: short-segment Hirschsprung disease (type I or S-HSCR), occurring in 60 to 85% of patients; and long-segment Hirschsprung disease (type II or L-HSCR), occurring in 15 to 25%. Loss-of-function mutations in RET and 13 other known, partially interdependent genes are involved. Different enhancer variants affect *RET* expression in a gene regulatory network involving a chromatin topologically associated domain as illustrated on p. 212. (Left and middle photographs: own observations; right photograph provided by Professor W. Hirsch, Leipzig.)

C. RET receptor

This is a transmembrane receptor encoded by the *RET* gene (164761) located at 10q11.2. It has 1,114 amino acids with 4 functional domains and is transcribed from 21 alternatively spliced exons. Mutations in HSCR are loss-of-function and are distributed along the gene. Mutations in *MEN2A* and *MEN2B* are gain-of-function and tend to cluster in different regions. *RET* mutations account for 15 to 35% of sporadic *HSCR* and 50% of familial cases. Three noncoding *RET* variants contribute to HSCR susceptibility.

D. RET and EDNRB receptor signaling

Three signaling pathways in the enteric nervous system are: (1) the RET receptor tyrosine kinase pathway and its ligand, glial cell line–derived neurotrophic factor (GDNF; 600837); (2) the endothelin type B receptor pathway (EDNRB; 131244) and its ligand, endothelin-3 (EDN3; 131242); and (3) the transcription factor SOX10 (602229). The resulting signals exert effects migration, and differentiation of neurons in the enteric ganglion cells.

For further details see table 7 in the appendix (p. 408).

Further Reading

Amiel J, et al.; Hirschsprung Disease Consortium. Hirschsprung disease, associated syndromes and genetics: a review. J Med Genet 2008; 45(1): 1–14

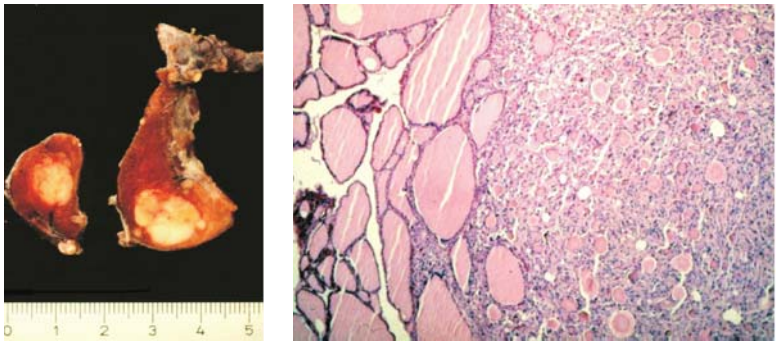
Chatterjee S, et al. Enhancer variants synergistically drive dysfunction of a gene regulatory network in Hirschsprung disease. Cell 2016; 167(2): 355–368.e10

Chatterjee S, et al. RET mutation and function in HSCR, MEN2 and other cancers. In: Erickson RP, et al., eds. Epstein's Inborn Errors of Development. 3rd ed. Oxford: Oxford Monographs of Genetics, 2016

Emison ES, et al. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. Nature 2005; 434(7035): 857–863

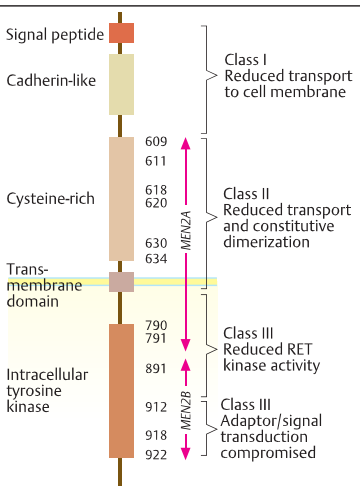
Emison ES, et al. Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. Am J Hum Genet 2010; 87(1): 60–74

Heanue TA, et al. Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. Nat Rev Neurosci 2007; 8(6): 466–479

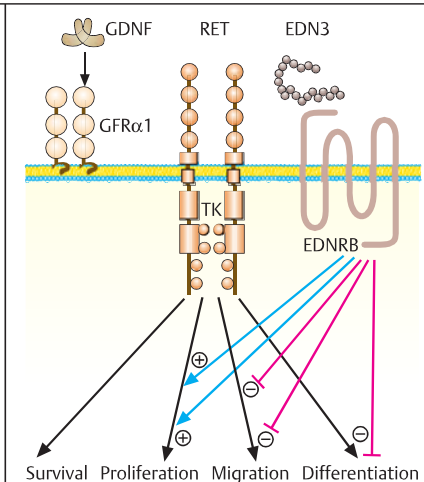


A. Multiple endocrine neoplasias

B. Congenital megacolon in Hirschsprung disease



C. RET receptor



D. RET and EDNRB signaling

Dysregulated RAS-MAPK Signaling Pathway

A clinically overlapping group of developmental disorders is caused by mutations in genes encoding components or modulators of the RAS-MAPK pathway. The RAS protein family (HRAS, KRAS, and NRAS; 190202) interacts via its effector RAF (164760) with mitogen-activated protein kinases (MAPK) in signaling pathways that regulate cell growth and differentiation. Noonan syndrome (163950), cardio-facio-cutaneous syndrome (CFC syndrome; 115150), Costello syndrome (218040), and LEOPARD syndrome (Noonan syndrome with multiple lentigines; 151100) are collectively called *RASopathies.*

A. Clinical phenotypes

Noonan syndrome (left) is the most common (1:2,500), and CFC syndrome (middle) and Costello syndrome (right) have characteristic facial features and are associated with congenital heart defects (predominantly pulmonary stenosis, hypertrophic cardiomyopathy, and septal defects), proportionate short stature and other developmental defects, variable cognitive deficits, and cancer predisposition.

B. The RAS-MAPK signaling cascade

The signaling cascade through RAF, MEK, and ERK (MAP kinases) is an important RAS-dependent pathway. It is initiated by binding of a growth factor to its receptor (receptor tyrosine kinase, RTK; see p. 186), followed by sequentially activated protein kinases. This eventually reaches intranuclear targets and regulates transcription. First, RTK is activated by phosphorylation. CBL (165360) is an adaptor protein of RTKs and regulates their degradation by ubiquitination (Ub). SOS1 (son of sevenless; 182530) acts as a guanine nucleotide exchange factor (GEF) recruited to activated RTKs. It activates RAS by exchanging GDP (guanosyldiphosphate) with GTP (guanosyltriphosphate). SHP2, a tyrosine phosphatase encoded by the gene *PTPN11* (176876), modulates signaling via the RAS-MAPK cascade. RAF kinases (BRAF, 164757; RAF1, 164760) become activated through binding to active GTP-bound RAS and subsequently phosphorylate MEK kinases (MEK1, 176872; MEK2, 601263), which in turn activate ERK (ERK1, ERK2). Phospho-ERK translocates to the nucleus where it phosphorylates nuclear targets. Neurofibromin (NF1) and SPRED1 (609291) are negative regulators in the RAS-MAPK cascade.

C. Mutations in RAS-MAPK pathway components

The RASopathies result from gain-of-function mutations in the genes encoding the signaling proteins, shown in **B**. Mutations in the gene *PTPN11* (**1**), protein-tyrosine phosphatase, nonreceptor-type 11, 176876, encoding SHP2, account for 50% of patients with Noonan syndrome. Most of the unevenly distributed mutations occur in different functional domains of the gene product SHP2. Germline mutations in the *KRAS* gene (**2**) account for approximately 2% of cases of Noonan syndrome and 5% of cases of CFC syndrome. The common somatic *KRAS* mutations in cancer differ and do not overlap despite clustering in similar regions. (Figures provided by Professor Martin Zenker, Magdeburg, Germany.)

D. Consequences of RAS mutations

Binding of GTP to RAS induces a conformational change in which two mobile regions (switch 1: yellow; switch 2: cyan) are involved. The inactive state is restituted through the interaction with GTPase-activating proteins (GAPs). Germline KRAS mutations occurring in Noonan and CFC syndromes (affecting residues G60 and P34) and the most common site of oncogenic mutations (G12) are located in this area and impair the RAS–GAP interaction, which results in RAS hyperactivation. (Figure courtesy of Professor M. R. Ahmadian, Düsseldorf, Germany.)

Further Reading

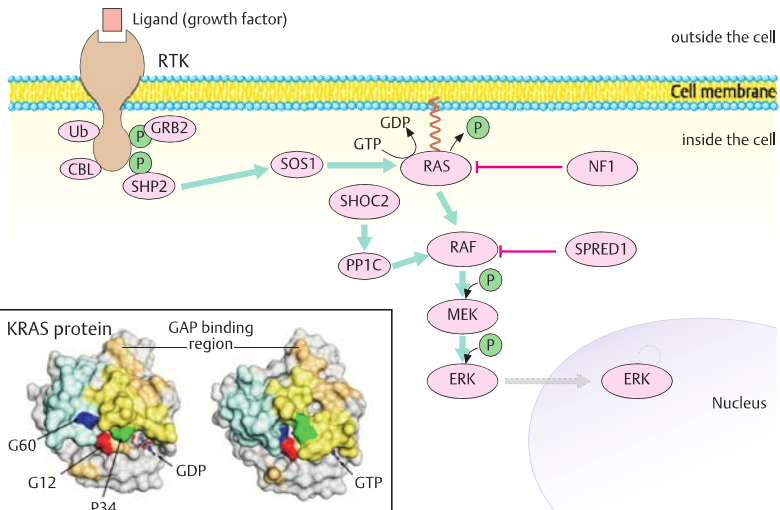
Gremer L, et al. Germline KRAS mutations cause aberrant biochemical and physical properties leading to developmental disorders. Hum Mutat 2011; 32(1): 33–43

Swensen J, et al. The Ras pathway. In: Erickson RP, et al., eds. Epstein's Inborn Errors of Development. 3rd ed. Oxford: Oxford University Press, 2016

Tartaglia M, et al. Noonan syndrome: clinical aspects and molecular pathogenesis. Mol Syndromol 2010; 1(1): 2–26

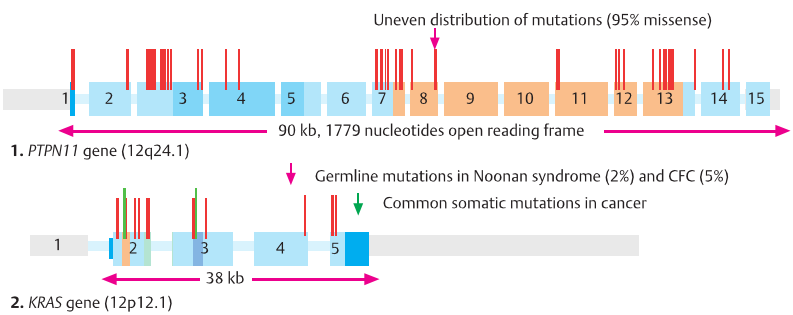
Zenker M, ed. Noonan Syndrome and Related Disorders. A Matter of Dysregulated Ras Signaling. Basel: Karger, 2009

A. Clinical phenotypes



B. The RAS-MAPK signaling cascade

D. RAS Mutations



C. Mutations in RAS-MAPK pathway components

Unstable Repeat Expansion

Expansion of unstable nucleotide repeats is the cause of approximately 30 neurogenic hereditary disorders (nucleotide expansion disorders). The expanded repeat is localized at a specific region near or within the gene involved (p. 84). In most cases, the repeat consists of three nucleotides. In others, a tetranucleotide or pentanucleotide repeat is expanded.

A. Types of disease

Unstable repeat diseases can be distinguished whether they are located in the untranslated regions of a gene (5′ UTR or 3′ UTR), in an exon, or in an intron. Functional consequences are loss of protein function or gain of an abnormal RNA level or protein, including loss of translational control (FRAXA), signaling (FRAXAE), or mitochondrial function (Friedreich ataxia, FRDA). In approximately half of these disorders, the codon CAG, encoding glutamine, is expanded within a coding region (polyglutamine disorders, or PolyQ), e.g., in Huntington disease.

B. Huntington disease

Huntington disease (HD, 143100), described in 1872 by George Huntington in a family from Long Island, occurs in approximately 4 to 7 individuals per 100,000. It is an autosomal dominant late-onset progressive neurodegenerative disease resulting from neuronal cell death and leading to complete loss of motor control and intellectual abilities within 5 to 10 years (**1**). It usually begins around the age of 40 to 50 years with uncoordinated movements (chorea, St Vitus' dance), excitation, hallucinations, and psychological changes. The 180-kb gene (*HTT*, 613004) is located on the distal short arm of human chromosome 4 at 4p16.3 between markers D4S127 and D4S125 (**2**), and has 67 exons. Two transcripts encode a ubiquitously expressed 3,144-amino-acid protein, huntington, involved in neuron function and survival. The 5′ coding region of the gene contains 4 to 35 copies of a trinucleotide repeat of cytosine, adenine, and guanine (CAG), which is a codon for glutamine. In patients, this is expanded to 40 to 250 CAG repeats. Incomplete penetrance may occur with 36 to 40 repeats. The CAG repeat length is inversely correlated with the age of onset. A diagnostic test distinguishes expanded and normal $(CAG)_n$ repeats (**3**). A predictive genetic test is possible many years prior to the onset of the first clinical signs. This requires prior genetic counseling according to established guidelines to ensure that informed consent of the individual tested has been achieved. (Image in 3, Zühlke et al., 1993, kindly provided by the late Professor W. Engel, Göttingen, Germany.)

C. Myotonic dystrophy

Myotonic dystrophy is an autosomal dominant neurological disease (**1**). A masklike face is characteristic (**2**). It occurs in two genetic types: DM1 (160900) caused by increased CTG repeats in the *DMPK* gene located at 19q13.32 (**3**) (605377); and DM2 (116955) caused by heterozygous expansion of a CCTG repeat in intron 1 of the zinc finger protein-9 gene (*ZNF9*, 602668; not shown). DM1 results from a pathogenic increase in the number of CTG repeats in the 3′ untranslated region beyond 50 repeats (**4**). The severity of the DM1 depends on the number of repeats. Southern blot analysis reveals increased sizes of DNA fragments (**4**). A neonatal form exists.

Further Reading

Gatchel JR, et al. Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet 2005; 6(10): 743–755

Harper P, et al. Myotonic dystrophy. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 5525–5550

Hayden MR, et al. Huntington disease. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 5677–5701

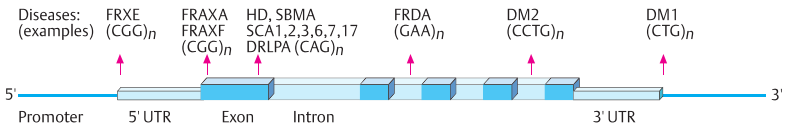
Keum JW, et al. The HTT CAG-expansion mutation determines age at death but not disease duration in Huntington disease. Am J Hum Genet 2016; 98 (2): 287–298

Orr HT, et al. Trinucleotide repeat disorders. Annu Rev Neurosci 2007; 30: 575–621

Walker FO. Huntington's disease. Lancet 2007; 369 (9557): 218–228

Zühlke C, et al. Mitotic stability and meiotic variability of the (CAG)n repeat in the Huntington disease gene. Hum Mol Genet 1993; 2(12): 2063–2067

Zühlke C, et al. Homozygous myotonic dystrophy: clinical findings in two patients and review of the literature. Am J Med Genet A 2007; 143A(17): 2058–2061

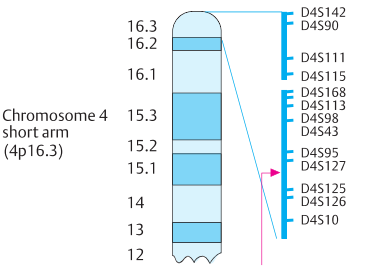


HD: Huntington disease (OMIM 143100)
SBMA: Spinobulbar muscular atrophy (OMIM 313200)
SCA: Spinocerebellar atrophies (OMIM 164400, 183090, 109150, 183086, 164500, 607136)
DRLPA: Dentatorubral-pallidolusian atrophy (OMIM 125370)
FRDA: Friedreich ataxia (OMIM 229300)
DM1: Myotonic dystrophy 1 (OMIM 160900); DM2 (OMIM 602668)

A. Types of disease

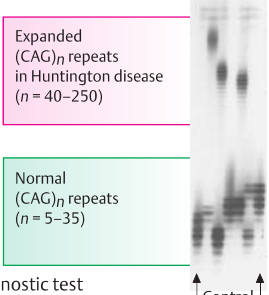
Severe progressive disease of the central nervous system
Loss of motor and intellectual functions
Onset age 25–60
Autosomal dominant
CAG repeat size increased
Predictive diagnosis possible, but problematic

1. Main manifestations



2. Localization of the gene

Affected individuals 1, 2, and 4 have expanded CAG repeats



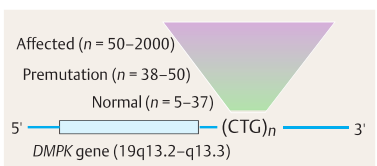
3. Diagnostic test

B. Huntington disease

Muscle weakness
Myotonia, mask-like face
Cataract, alopecia
Variable expression
Autosomal dominant
CTG repeat increased

1. Main manifestations

2. Phenotype



3. Expanded CTG repeat in myotonic dystrophy

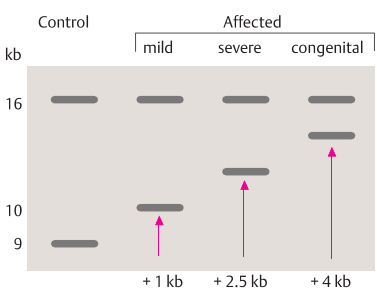


Diagram of a Southern blot at gene locus D19S95 (probe pBB0.7)

4. Correlation with degree of severity

C. Myotonic dystrophy

Fragile X Syndrome

Fragile X syndrome (300624; synonyms: fragile X mental retardation syndrome, FMR1, or Martin–Bell syndrome) is a form of mental retardation with a prevalence of approximately 1:4,000 males and 1:6,000 females. It is caused by expansion of a CGG trinucleotide repeat in the 5′ untranslated region of the *FMR1* gene (309550) on the distal long arm of the X chromosome at Xq27.3. It was identified in 1991 by Oberle and Verkerk and coworkers as the first human disease resulting from expansion of an unstable trinucleotide repeat sequence. Related disorders are FraX-associated tremor/ataxia syndrome (300623) and FraX-associated premature ovarian insufficiency (311360).

A. Phenotype

Individuals with fragile X syndrome have varying degrees of intellectual developmental delay associated with behavioral and physical features. Some patients have connective tissue weakness. The testes are usually enlarged.

B. Fragile site FRAXA

This disorder derived its name from a cytogenetically visible fragile site at region Xq27 of the X chromosome in lymphocytes cultured in medium deficient in folic acid. Fragile X tremor/ataxia syndrome (FXTAS; OMIM 300623) results from an abnormal gain-of-function of FMR1 RNA. Other variants with different fragile sites exist, for example, FRAXE at Xq28 (OMIM 309548).

C. *FMR1* gene and protein

An inactivated *FMR1* gene due to an increased number of more than 200 CGG repeats (full mutation) in the 5′ upstream region of the gene is by far the most common cause of this disorder. The high content of GC induces aberrant methylation and inhibits transcription with loss of the FMR1 protein and defective translational control of neuronal synaptic proteins. The normal number of CGG repeats is 6 to 50. A premutation predisposing to further expansion has 59 to 200 repeats. The *FMR1* gene has 17 exons spanning 38 kb. Its transcript is alternatively spliced and translated into at least 20 protein isoforms (FMRPs). The protein (FMRP) selectively binds RNA. At least two functional domains are RNA-binding sites, KH2 and RGG (KH domains consist of 40–60 amino acids with invariant hydrophobic leucine, isoleucine, or methionine; RGG consists of a 20–30 amino acid motif with arginine–glycine–glycine residues). In Drosophila, *Fmr1* is a component of the RNA-induced silencing complex (RISC; see p. 174).

D. Inheritance and genetic testing

Southern blot analysis (**1**) can distinguish individuals with fragile X syndrome (Fra X) carrying a full mutation allele (>200 CGG trinucleotides), a premutation (59–200), or a normal allele (6–50). The visible bands (arrows) are DNA fragments harboring the CGG repeat region, derived from genomic DNA cleaved by restriction endonuclease PstI, and hybridized to the radiolabeled *FMR1* probe Ox0.55. Polymerase chain reaction (PCR) protocols determine the lengths of normal and premutation alleles. Expansion leads to hypermethylation and transcriptional silencing of the *FMR1* gene. (Figure kindly provided by Professor P. Steinbach, Ulm, Germany.) Female carriers of a permutation inherit this usually from their father. Daughters of male carriers inherit an allele with a permutation. In the pedigree (**2**), the number of CGG repeats at the *FMR1* locus is shown for each individual. A premutation can be transmitted by a female (I-2, II-3, III-2) or a male (II-2). The premutation allele may expand into a full mutation when passed from a mother to her children.

Further Reading

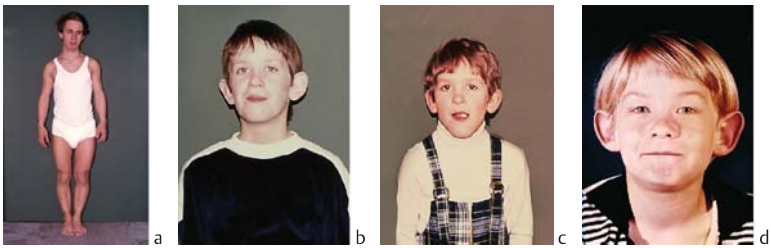
Hagerman R, et al. Advances in clinical and molecular understanding of the FMR1 premutation and fragile X-associated tremor/ataxia syndrome. Lancet Neurol 2013; 12(8): 786–798

Oberlé I, et al. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 1991; 252(5009): 1097–1102

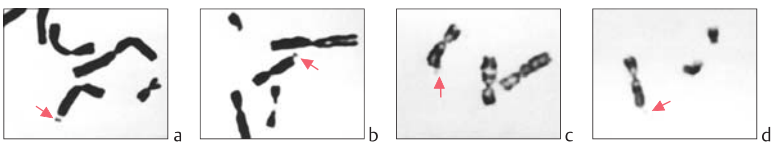
Penagarikano O, et al. The pathophysiology of fragile X syndrome. Annu Rev Genomics Hum Genet 2007; 8: 109–129

Verkerk AJ, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 1991; 65(5): 905–914

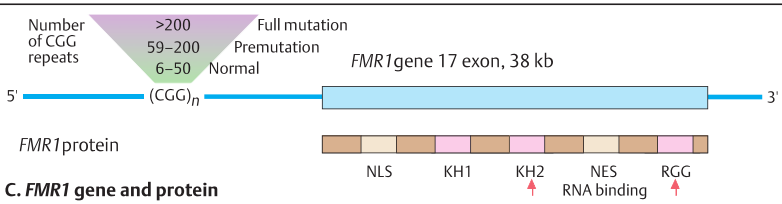
Warren ST, et al. The fragile X syndrome. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 1257–1289



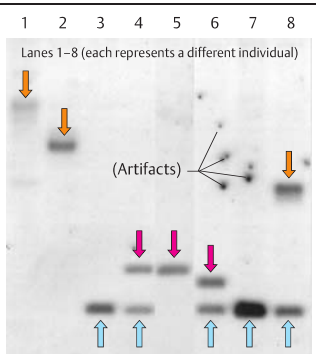
A. Phenotype



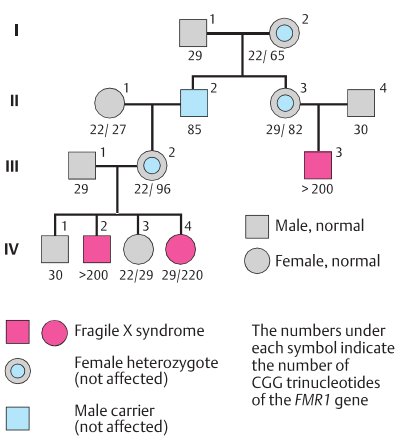
B. Fragile site Xq27.3



C. *FMR1* gene and protein



1. Southern blot analysis for Fra X syndrome



2. Phenotypic effects of expanded CGG repeats

D. Inheritance and genetic testing

Imprinting Disorders

Imprinting disorders result from various defects resulting in faulty imprinting of the particular chromosomal region involved (p. 182). Different mechanisms affect one or more active genes normally expressed in only one parental allele in an imprinted region. Best known are Prader–Willi syndrome (PWS; OMIM 176270) and Angelman syndrome (AS; 105830) at region 15q11–13, and Beckwith–Wiedemann syndrome (130650) at 11p15.5.

A. Prader–Willi and Angelman syndromes

PWS and AS are neurogenetic developmental disorders resulting from different genetic lesions in an imprinted region of human chromosome 15 (15q11–13). It results in PWS if the chromosome 15 involved is of paternal origin, but in AS if the lesion involves the chromosome 15 of maternal origin. PWS is characterized by neonatal muscular weakness and feeding difficulties, followed in early childhood by reduced or lack of satiation control, which leads to massive obesity in many patients. In AS, the developmental retardation is usually severe, with nearly complete lack of speech development, an abnormal electroencephalogram, tendency to seizures, and hyperactivity.

B. Imprinted Chromosomal Region 15q11-q13

The simplified map of the imprinted chromosomal region 15q11-q13 consists of gene loci expressed from the allele of paternal origin (pat) only, as shown in blue. Under the control of an imprinting center (IC), these are transcribed in the same or the opposite direction. The corresponding gene loci of maternal origin (mat) are permanently silenced due to DNA methylation. On the other hand, a gene, UBE3A (ubiquitin-protein ligase E3, 601623) at 15q11.2 is transcribed from the mat allele, whereas the pat counterpart of it is silenced. The IC, controlling the entire imprinted region, appears to consist of two elements; one is required for the maintenance of the paternal imprint during early embryogenesis, whereas the other controls the maternal imprinting in the female germline.

C. Prader-Willi Syndrome

Prader–Willi syndrome (PWS, 176270) results from the loss of expression of the imprinted paternal region comprising SNRPN gene (182279), NDN gene (602117), and other genes in the region 15q11-q13. An interstitial deletion of these PWS-critical genes the most frequent cause of PWS in 70% of patients, since only the inactive maternal (mat) alleles are present (1). Maternal uniparental disomy (matUPD) is responsible for PWS in about 25–30% of patients. The presence of two identical maternal chromosomes leads to a condition known as isodisomy; if these differ, it is referred to as heterodisomy. Different defects involving the imprinting center (IC) are rare but important causes. If the breakpoint of a chromosome translocation is located in this imprinted region, it may also result in PWS.

D. Angelman Syndrome

Angelman syndrome (AS, 105830) is caused if the lesion involves the AS-critical region with maternal expression of genes in this region. The gene affected is UBE3A (ubiquitin-protein ligase E3, 601623) at 15q11.2. It is characterized by maternal deletion in approximately 70%, paternal uniparental disomy (UPD) in approximately 2%, and an imprinting error in approximately 1% patients. In approximately 15%, the cause remains unknown. Defects in the imprinting center (IC) and mutations in the UBE3A gene may lead to familial occurrence of PWS or AS (data for the figures in B, C, and D are provided by B. Horsthemke, Essen).

Further Reading

Buiting K. Prader-Willi syndrome and Angelman syndrome. Am J Med Genet C Semin Med Genet 2010; 154C(3): 365–376

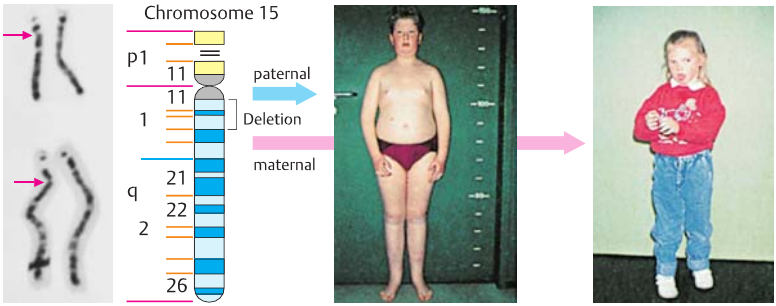
Horsthemke B. Mechanisms of imprint dysregulation. Am J Med Genet C Semin Med Genet 2010; 154C(3): 321–328

Horsthemke B, et al. Imprinting mutations on human chromosome 15. Hum Mutat 1997; 10(5): 329–337

Horsthemke B, et al. Imprinting defects on human chromosome 15. Cytogenet Genome Res 2006; 113(1–4): 292–299

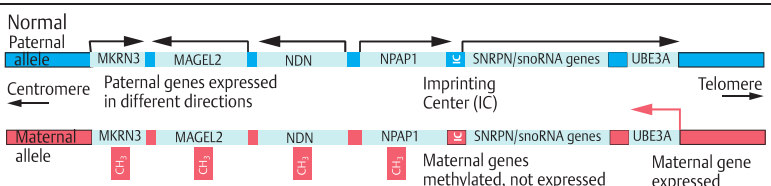
Horsthemke B, Wagstaff J. Mechanisms of imprinting of the Prader-Willi/Angelman region. Am J Med Genet A 2008; 146A(16): 2041–2052

Larson AM, et al. Angelman syndrome in adulthood. Am J Med Genet A 2015; 167A(2): 331–344

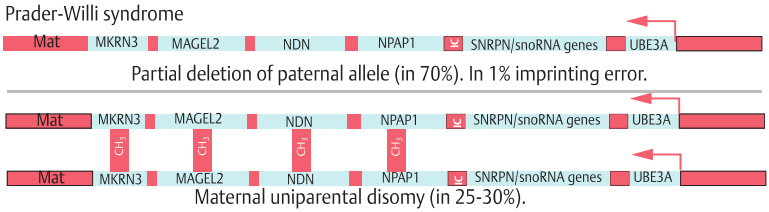


1. Interstitial deletion 15q11-13 **2.** Prader–Willi syndrome **3.** Angelman syndrome

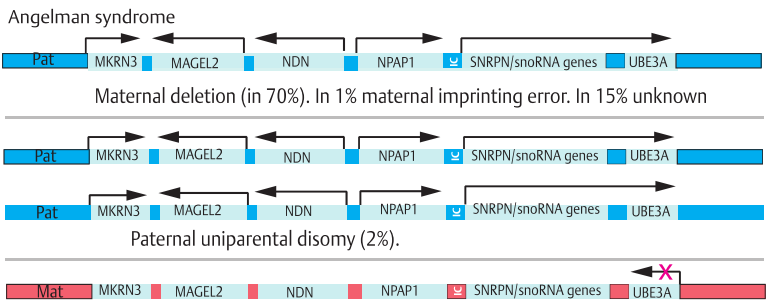
A. Prader-Willi and Angelman syndromes



B. Imprinted chromosomal region 15q11-q13



C. Genetic causes of Prader-Willi syndrome



D. Genetic causes of Angelman syndrome

Mitochondrial Diseases

Human mitochondrial dysfunction causes a wide variety of different degenerative and acute disorders. It also is involved in aging and carcinogenesis. Mutations and deletions occur in human mitochondrial DNA (mtDNA), and they can be either inherited or de novo. Mitochondrial mutations are transmitted by maternal inheritance only. However, mutations in nuclear genes may influence mitochondrial function and manifest as monogenic inheritance.

The clinical spectrum and age of onset of mitochondrial diseases vary widely. The age of onset and severity are highly variable, even within a family affected by the same mutation. Organs with high-energy requirements are particularly vulnerable: the brain, heart, skeletal muscle, eye, ear, liver, pancreas, and kidney. These organs are affected by several neurodegenerative and cardiovascular disorders, diabetes mellitus, deafness, and cancer. Acquired mitochondrial mutations accumulate with age.

The mutation rate of mtDNA is 10 times higher than that of nuclear DNA. Mitochondrial mutations are generated during oxidative phosphorylation through pathways involving reactive oxygen molecules. Mutations accumulate because effective DNA repair and protective histones are lacking. At birth, most mtDNA molecules are identical (*homoplasmy*); later, they differ as a result of mutations accumulated in different mitochondria (*heteroplasmy*).

A. Mutations and deletions in human mitochondrial DNA

Mitochondrial genetic disorders result from deletions, point mutations, and other structural rearrangements. Some are characteristic and recur in different, unrelated patients. The extent of three common deletions of 10.4, 7.4, and 5 kb are shown as partial circles outside the mtDNA map. Inside the site of common mutations are shown: LHON (Leber's optical nerve atrophy; 535000) at nucleotide positions 11778 and 3460; NARP (neuropathy, ataxia, retinitis pigmentosa; 551500) at 8993; MERRF (myoclonus epilepsy with ragged red fibers in muscle cells; 545000) at 8344; MMC (mitochondrial myopathy and cardiomyopathy; 590050) at 3260; MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis; 540000) at 3243; and a mutation A to G at 1555, which confers high sensitivity to aminoglycosides, causing deafness (see Table 8 on p. 409). (Figure adapted from Wallace et al., 2007, and MITOMAP.)

B. Maternal inheritance of a mitochondrial disease

Both somatic and inherited mutations occur. Hereditary mitochondrial diseases are transmitted through the maternal line only, since spermatozoa contain very few, if any, mitochondria. Thus, the disease will not be transmitted from an affected man to his children.

C. Heteroplasmy for mitochondrial mutations

Many mutations or deletions in mitochondria are acquired during the lifetime of an individual. Their proportion may be different in different tissues and increase with age. This difference is referred to as heteroplasmy. It contributes to the considerable variability of mitochondrial diseases. A germline mutation may be present in all cells (homoplasmy). The proportion of defective mitochondria varies after repeated cell divisions.

Recently, replacement of defective mitochondria by a donor egg cell has been attempted (Kang et al, 2016).

Further Reading

Genetics Home Reference. Mitochondrial DNA. Available at: http://ghr.nlm.nih.gov/chromosome/MT. Accessed February 26, 2017

Kang E et al. Mitochondrial replacement in human oocytes carrying pathogeneic mitochondrial DNA mutations. Nature 2016; 540: 270–275

Koopman WJ, et al. Monogenic mitochondrial disorders. N Engl J Med 2012; 366(12): 1132–1141

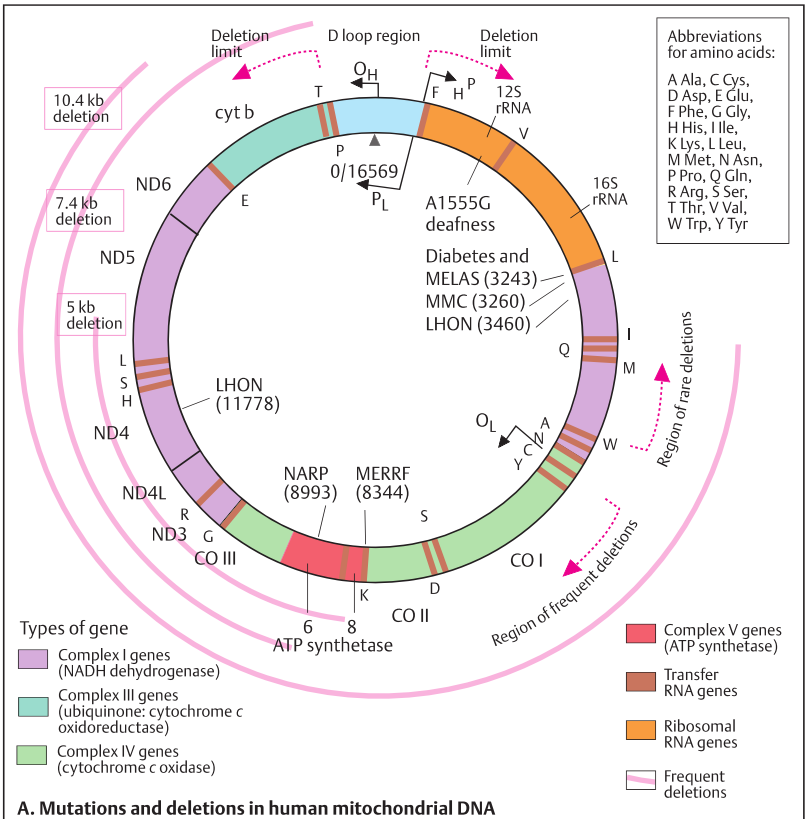
Lightowlers RN, et al. Mutations causing mitochondrial disease: what is new and what challenges remain? Science 2015; 349(6255): 1494–1499

MITOMAP. A human mitochondrial genome database. Available at: www.mitomap.org/MITOMAP

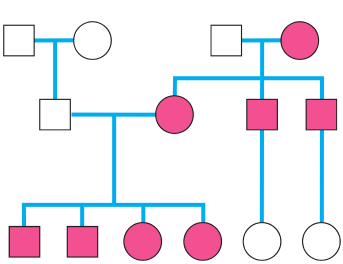
Schon EA, et al. Human mitochondrial DNA: roles of inherited and somatic mutations. Nat Rev Genet 2012; 13(12): 878–890

Wallace DC. Genetics: mitochondrial DNA in evolution and disease. Nature 2016; 535(7613): 498–500

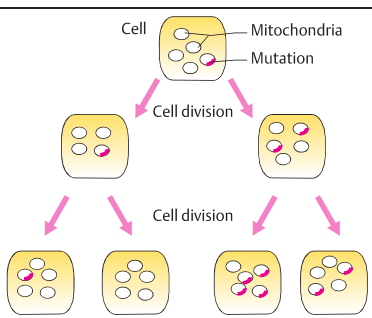
Wallace DC, et al. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. Cold Spring Harb Perspect Biol 2013; 5 (11): a021220



A. Mutations and deletions in human mitochondrial DNA



B. Maternal inheritance of a mitochondrial disease



C. Heteroplasmy for mitochondrial mutations

Chloride Channel Defects: Cystic Fibrosis

Cystic fibrosis (CF, mucoviscidosis; 219700) is a highly variable multisystem disorder caused by mutations in the cystic fibrosis transmembrane conduction regulator gene, encoding a chloride channel protein (*CFTR*; 602421). It is one of the most frequent autosomal recessive hereditary diseases in populations of European origin (approximately 1 in 2,500 newborns). The high frequency of heterozygotes (1:25) is thought to result from a selective advantage: heterozygotes may be less susceptible to cholera than normal homozygotes. Polymorphisms in other genes modify the severity of pulmonary manifestation. Examples of modifying genes are the *TGF1* gene on 19q13.1 (190180) and *IFRD1* (603502).

A. Cystic fibrosis: clinical aspects

The disease primarily affects the bronchial system and the gastrointestinal tract. Viscous mucus formation leads to frequent, recurrent lung and bronchial infections, resulting in chronic pulmonary insufficiency. The average life expectancy in typical CF has been increased in recent years, but still is approximately 30 years. The disease may take a less severe, almost mild, course. Certain mutations occur in males with bilateral absence of the vas deferens (CBAVD, 277180).

B. Finding the gene for cystic fibrosis

Finding the gene is the best example to illustrate how genes had to be identified prior to the direct sequencing now available. It was one of the first human genetic diseases identified by chromosome walking and jumping (positional cloning). Linkage analysis first localized the gene to chromosome 7 at q31.2. A long-range restriction map of approximately 1,500 kb was then narrowed down to 250 kb, and the gene was identified and characterized by showing homozygous mutations in patients that were absent in controls.

C. The *CFTR* gene and its protein

The large *CFTR* gene spans over 250 kb of genomic DNA, organized into 27 exons (exons 6 and 14 are, respectively, numbered 6a/6b and 14a/14b), which encode a 6.5-kb transcript with several alternatively spliced forms of mRNA. The protein of 1,480 amino acids is a membrane-bound chloride ion channel regulator with 5 functional domains, as shown. Two nucleotide-binding domains 1 and 2 (NBD1 and NBD2) confer cAMP-regulated chloride channel activity. The most common mutation is a deletion of a phenylalanine codon in position 508 (F508del). The regulatory R domain contains numerous sites for phosphorylation. *CFTR* is widely expressed in epithelial cells.

More than 2000 mutations observed in the *CFTR* gene are classified according to: (1) abolished synthesis of full-length protein, (2) block in protein processing, (3) reduced chloride channel regulation, (4) reduced chloride channel conductance, and (5) reduced amount of normal CFTR protein. The underlying genetic defects include missense mutations (45%), nonsense mutations (10%), RNA splice mutations (13%), small deletions/insertions (21%), large deletions (6%), and regulatory sequences (1.4%). The most frequent mutation, a deletion of phenylalanine codon 508 (ΔF508, now named F508del), accounts for approximately 66% of patients in Europe. Other mutations above 1% are p.(Asn1303Lys), p.(Arg553X), p.(Gly551Asp), and p.(Arg34Pro). They are included in molecular testing along with other mutations.

To some extent, the type of mutation predicts the severity of the disease. The most severe forms are associated with homozygosity for ΔF508 and compound heterozygotes ΔF508/G551D and ΔF508/G542X.

Further Reading

Brennan ML, et al. Cystic fibrosis: a review of associated phenotypes, use of molecular diagnostic approaches, genetic characteristics, progress, and dilemmas. J Mol Diagn 2016; 18(1): 3–14

Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat Rev Genet 2015; 16(1): 45–56

Cystic Fibrosis Mutation Database. Hospital SickKids, Toronto, Canada. Available at: http://www.sickkids.ca/Centres/Cystic-Fibrosis-Centre/. Accessed March 14, 2017

MacKenzie T, et al. Longevity of patients with cystic fibrosis in 2000 to 2010 and beyond: survival analysis of the Cystic Fibrosis Foundation patient registry. Ann Intern Med 2014; 161(4): 233–241

Saint-Criq V, Gray MA. Role of CFTR in epithelial physiology. Cell Mol Life Sci 2017; 74: 93–115

Cystic fibrosis (mucoviscidosis)

Severe progressive disease of the bronchial system and gastrointestinal tract

Disturbed function of a chloride ion channel by mutations in the *CFTR* gene

Autosomal recessive

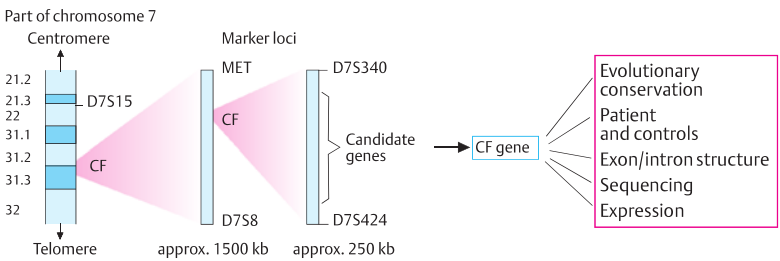
Gene locus 7q31.2

Disease incidence approx. 1:2500

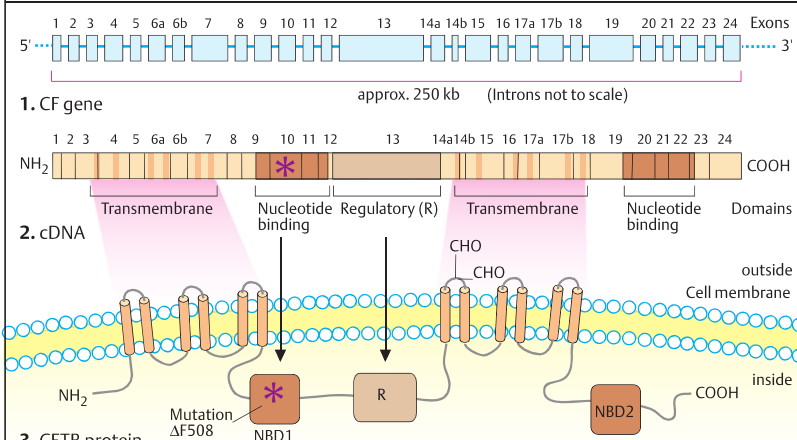
Heterozygote frequency approx. 1:25

Mutation ΔF508 in approx. 70%

A. Cystic fibrosis: clinical aspects



B. Finding the gene for cystic fibrosis (CF)



C. The *CFTR* gene and its protein

Genetic Defects in Ion Channels: LQT Syndromes

Mutations in genes encoding ion channels cause more than 30 different monogenic disorders (channelopathies). Ion channels play a prominent physiological role in the functioning of cardiac and striated muscle, neuromuscular junctions, and neurons of the central nervous system, inner ear, and retina. A striking example is the long QT (LQT) syndrome, which comprises 14 genetically different disorders: LQT syndrome types 1 to 14 (see OMIM 192500), caused by mutations in genes encoding different types of ion channels or their interaction partners. The main manifestation is cardiac arrhythmia leading to cardiac arrest. It is important to distinguish the different types, because the choice of medication differs. Approximately 10% of patients carry a second mutation in the same or another ion channel gene.

A. Long QT syndrome, a genetic cardiac arrhythmia

The main features of LQT syndrome (**1**) are characterized by a prolonged QT interval (**2**) in the electrocardiogram (>440 ms, corrected for heart rate). Other diagnostic features are aberrant T-wave morphology and torsade de pointes ventricular tachycardia (rapid heartbeats). It consists of 14 genetic types (11 shown).

B. Different molecular types of long QT syndrome

Prolongation of the QT interval in the electrocardiogram results from an increased duration of the cardiac action potential (**1**). In phase 0, the myocardiocyte is depolarized by sodium currents ($Na_v1.5$, encoded by the *SCN5A* gene). In plateau phases 1 and 2, the normal potential is supported by calcium currents ($Ca_v1.2$, *CACNA1C*). In phase 3, the resting membrane potential is restored by potassium currents ($K_v7.1$, *KCNQ1*, and $K_v11.1$, *KCNH2*). LQT1 (**2**, caused by mutations in the *KCNQ1* gene encoding a potassium channel) accounts for approximately 30% of patients with long QT syndrome. LQT2 (**3**, mutations in the *KCNH2* gene [formerly *HERG*]) account for 10% of patients. *KCNH2* encodes a 1,159-amino-acid transmembrane protein of the other major potassium channel that participates in phase 3 repolarization. LQT3 (**4**, *SCN5A*, the gene encoding the main sodium channel protein Na1.5) accounts for 3% of patients. It consists of four subunits (I–IV), each containing six transmembrane domains and several phosphate-binding sites. Other, rare, LQT syndromes are caused by increased sodium currents (LQT3, LQT9, LQT10, and LQT12), augmented calcium currents (LQT4 and LQT8) or decreased potassium currents (LQT1, LQT2, LQT5, LQT6, LQT7, LQT11, and LQT13).

The type of mutation influences the clinical course and the response to therapeutic drugs.

Syndromic forms of LQT include: the autosomal recessive Jervell and Lange-Nielsen syndrome (OMIM 220400) associated with hearing defects, and caused by homozygous mutations in *KVLQT1* or *KCNE1* (LQT5) genes encoding the α and β subunits of the IKS channel; the autosomal dominant Romano–Ward syndrome (OMIM 192500) caused by mutations in *KVLQT1*, the LQT1 gene; and other syndromic forms (see Appendix, Table 9, p. 409). (Figure adapted from Ackerman and Clapham, 1997.)

Further Reading

Ackerman MJ. Cardiac channelopathies: it's in the genes. Nat Med 2004; 10(5): 463–464

Ackerman MJ, et al. Ion channels--basic science and clinical disease. N Engl J Med 1997; 336(22): 1575–1586

Beckmann BM, et al. Inherited cardiac arrhythmias: diagnosis, treatment, and prevention. Dtsch Arztebl Int 2011; 108(37): 623–633

Keating MT, et al. Molecular and cellular mechanisms of cardiac arrhythmias. Cell 2001; 104(4): 569–580

Marks AR. Arrhythmias of the heart: beyond ion channels. Nat Med 2003; 9(3): 263–264

Modell SM, Lehmann MH. The long QT syndrome family of cardiac ion channelopathies: a HuGE review. Genet Med 2006; 8(3): 143–155

Roden DM. Clinical practice. Long-QT syndrome. N Engl J Med 2008; 358(2): 169–176

Schwartz PJ et al. Long-QT syndrome. Circulation: Arrhythmia adnd Elctrophyiolol 2012;5: 858–877

Long QT Syndrome. Medscape at http://emedicine.medscape.com/article/157826-overview (accessed 16 August 2017)

Zareba W, et al. International Long-QT Syndrome Registry Research Group. Influence of the genotype on the clinical course of the long-QT syndrome. N Engl J Med 1998; 339(14): 960–965

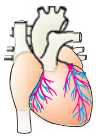
Prolonged QT interval in the electrocardiogram

Syncope

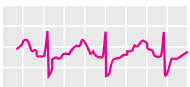
Sudden death

Autosomal dominant

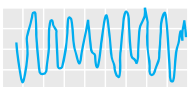
13 genes involved (LQT1 - LQT8)



Prolonged QT



Torsade de pointes

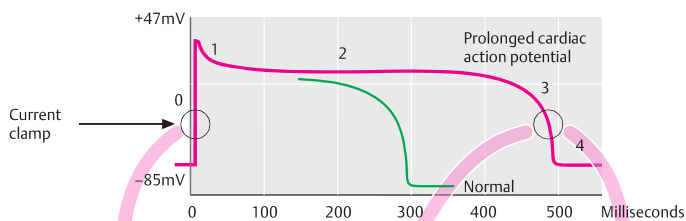


Long QT syndromes

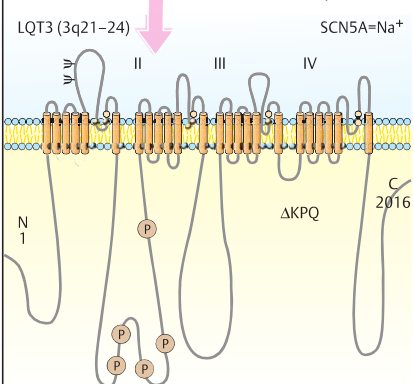
Type	Locus	Gene
LQT1	11p15.5	*KCNQ1 (KVLQT1)*
LQT2	7q35–36	*HERG*
LQT3	3p21–24	*SCNA5*
LQT4	4q25–27	*Ankyrin-B*
LQT5	21q22.1	*KCNE1*
LQT6	21q21.1	*KCNE2*
LQT7	17q23	*KCNJ2*
LQT8	12p13.2	*CACNA1c*
LQT9	3p25	*Caveolin-3*
LQT10	11q23	*SCN4B*
LQT11	7p21–q22	*AKAP9*

1. Main features **2.** Electrocardiogram **3.** Genetics

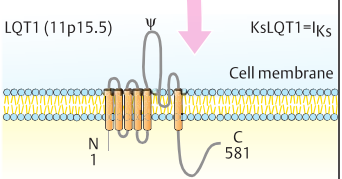
A. Long QT syndrome, a genetic cardiac arrhythmia



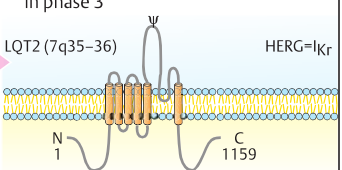
1. Increased duration of cardiac action potential



4. Na^+-channel fails to inactivate completely during phase 0



2. Voltage-activated K^+-channel delayed in phase 3



3. Voltage-gated K^+-channel delayed in phase 3

B. Different molecular types of long QT syndrome

α_1-Antitrypsin Deficiency

α_1-Antitrypsin deficiency (A1ATD, 613490) is an autosomal recessive disorder of lung function. α_1-Antitrypsin is the main protease inhibitor in blood plasma, first described by Laurell and Eriksson (1963). As a member of the serine protease inhibitor superfamily (SERPIN), it has a major role in activating neutrophil elastase and other proteases to maintain protease/antiprotease balance. It binds to a wide range of proteases, such as elastase, trypsin, chymotrypsin, thrombin, and bacterial proteases. Its most important physiological effect is the inhibition of leukocyte elastase in the bronchial system.

A. α_1-Antitrypsin

Human α_1-antitrypsin is a 52-kDa glycoprotein composed of 394 amino acids and 12% carbohydrate content. It is encoded by the *SERPINA1* gene (107400) located at 14q32.13 (genomic coordinates: *14:94,376,746–94,390,691*).

B. α_1-Antitrypsin deficiency

α_1-Antitrypsin deficiency is a variable disease with chronic obstructive pulmonary emphysema as the main sign. It usually manifests by the third or fourth decade of life due to uninhibited activity of leukocyte elastase on the elastin of the pulmonary alveoli. This is visible by increased darkness of the chest X-ray. The concentration of α_1-antitrypsin (α_1-AT) in alveolar fluid is greatly reduced in heterozygotes (red squares) and more so in homozygotes (yellow triangles). This can be corrected by intravenous administration of α_1-antitrypsin. Since oxidizing substances have an inhibitory effect, smokers have a much more rapid clinical course (onset of dyspnea at 35 years of age instead of 45–50 years). A related, allelic disorder is α_1-antichemotrypsin deficiency (107280). (Figure kindly provided by N. Konietzko, Essen.)

C. α_1-Antitrypsin mutations

More than 100 allelic variants are known. The α_1-antitrypsin protein is highly polymorphic due to differences in the amino acid sequence and variations in the three carbohydrate side chains at positions 46, 83, and 247. The main clinically important mutations occur at the sites marked by red arrows. The *SERPINA1* gene (107400) spans 12.2-kb DNA with a 1,434-bp coding region in five exons on chromosome 14 (14q32.13). The reactive site is located at position 358/359 (methionine/serine). The α_1-antitrypsin alleles are grouped into four classes: (1) normal, (2) deficiency, (3) null alleles, and (4) dysfunctional alleles. The typical normal allele is Pi(M); the most important deficiency alleles are Pi(Z), Pi(P), and Pi(S). The most frequent deficiency allele, Pi(Z), causes plasma concentrations of α_1-antitrypsin of approximately 12 to 15% of normal in the homozygous genotype Pi(ZZ) and 64% in the heterozygote (Pi*MZ). MS heterozygotes have 86% of the MM homozygote activity. The heterozygote frequency estimated by Lazarin and coworkers in European populations is approximately 1 in 10, compared to 1 in 249 in Asia.

D. Synthesis of α_1-antitrypsin

The α_1-AT gene (*SERPINA1*) is expressed in liver cells (hepatocytes). The gene product is channeled through the Golgi apparatus and released from the cell (secreted). The Z mutation leads to aggregation of the enzyme in the liver cells, with too little of the enzyme being secreted. The S mutation leads to premature degradation.

E. Reactive center of protease inhibitors

As a member of a family of protease inhibitors, α_1-AT shows marked homology with related protein, especially at their reactive centers. Oxidizing substances have an inhibitory effect and inactivate the molecule. (Figures adapted from Cox, 2001.)

Further Reading

Cox DW. α_1-Antitrypsin deficiency. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill; 2001: 5559–5584

Hubbard RC, et al. Biochemical efficacy and safety of monthly augmentation therapy for alpha 1-antitrypsin deficiency. JAMA 1988; 260(9): 1259–1264

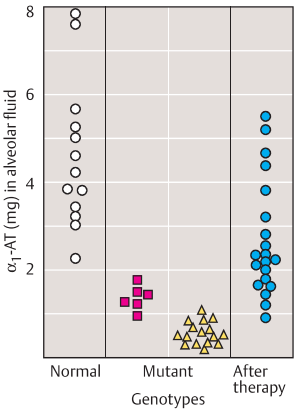
Laurell C-B, Erickson S. The electrophoretic alpha-1-globulin pattern of serum in alpha-1-antitrypsin deficiency. Scand J Clin Lab Invest 1963; 15: 132–140

Lazarin GA, et al. An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals. Genet Med 2013; 15(3): 178–186

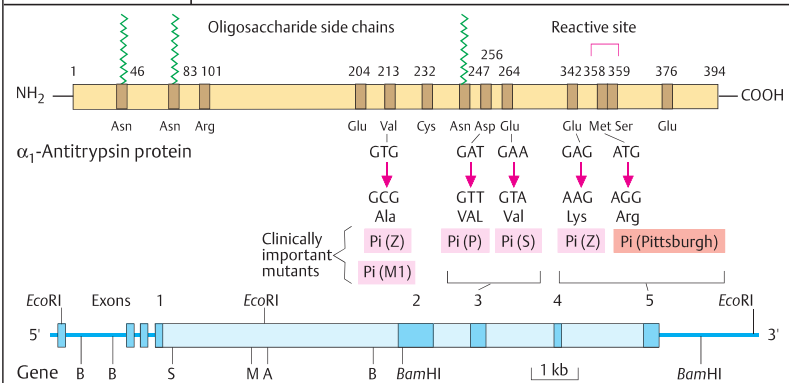
Stoller JK, Aboussouan LS. α_1-antitrypsin deficiency. Lancet 2005; 365(9478): 2225–2236

Function:
main plasma protease inhibitor, especially of elastase in the bronchial system of the lung
Gene product:
52 kDa glycoprotein
12% carbohydrate content
Gene:
10.2 kb gene,
5 exons
gene locus 14q32.1

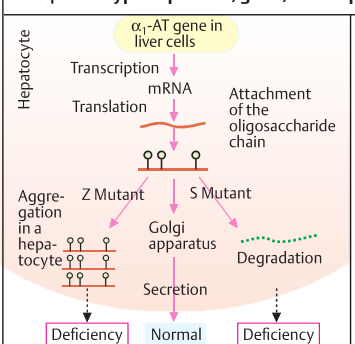
A. α1-Antitrypsin



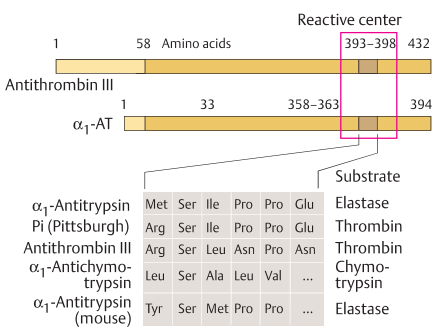
B. α1-Antitrypsin deficiency



C. α1-Antitrypsin: protein, gene, and important mutants



D. Synthesis of α1-antitrypsin



						Substrate	
α1-Antitrypsin	Met	Ser	Ile	Pro	Pro	Glu	Elastase
Pi (Pittsburgh)	Arg	Ser	Ile	Pro	Pro	Glu	Thrombin
Antithrombin III	Arg	Ser	Leu	Asn	Pro	Asn	Thrombin
α1-Antichymo-trypsin	Leu	Ser	Ala	Leu	Val	...	Chymo-trypsin
α1-Antitrypsin (mouse)	Tyr	Ser	Met	Pro	Pro	...	Elastase

E. Reactive center of protease inhibitors

Hemophilia A and B

Hemophilia A is a severe bleeding disorder resulting from inactivity of blood coagulation factor VIII (306700) due to X-chromosomal mutations in the gene F8 (300841). A more rare (1:30,000) related disorder is hemophilia B (306900) resulting from X-chromosomal deficiency of factor IX (F9). Factors VIII and IX function in the blood coagulation cascade. Factor VIII functions as a cofactor in the activation of factor X to factor Xa. Hemophilia was the first recognized major disease to be determined to have a genetic cause. The Talmud refers to its increased occurrence in males in certain families. F. Hopff (Würzburg, Germany) introduced the term hemophilia in a medical thesis in 1828.

A. Inheritance of hemophilia

The X-chromosomal inheritance is readily apparent in some royal families in Europe. Several descendants of Queen Victoria, a heterozygous carrier, inherited hemophilia. Hemophilia A occurs with a frequency of approximately 1 in 5,000 males. Approximately 30% of cases result from a new mutation. "Royal hemophilia" is now recognized to be hemophilia B (Rogaev et al., 2009).

B. Blood coagulation factor VIII

Activated factor VIII protein (by thrombin) consists of five subunits (A1, A2, A3, C1, C2) held together by calcium ions (Ca^{2+}; **1**). The inactive factor VIII protein (**2**) contains three domains (A, B, C). Domain A consists of three homologous copies (A1, A2, A3), domain C of two copies (C1, C2), and domain B of one copy. In humans, the F8 gene (**3**) is located at Xq28, near the gene for factor IX. It consists of 26 exons and spans 186 kb of DNA. Noteworthy are the large exon 14 (3106 bp), which encodes the B domain, and the large intron 22 (32,000 bp) between exons 22 and 23. Most mutations involve the dinucleotide CG in DNA sequences TCGA. This mutates easily to TTGA because the cytosine of this dinucleotide is frequently methylated. Subsequent deamination of methylcytosine leads to a C to T transition. This creates a stop codon (TGA), resulting in a truncated factor VIII protein.

Genetic tests employ a targeted mutation search or sequencing of the entire coding region. RFLP (restriction fragment length polymorphisms) analysis has been utilized (**4**). A polymorphic variant in the recognition sequence for the restriction enzyme *BclI* near exons 17 and 18 produces fragments of 879 bp and 286 bp; when the variant is absent, a single fragment of 1,165 bp is produced. In the family pedigree (**5**), two affected males (II-1 and III-2) carry the 879-bp fragment. Thus, this fragment indicates presence of the mutation.

Common types of mutation in hemophilia A are nonsense in 14% of patients, small deletions and insertions in 15%, splice mutations in 4%, and inversion flip tip mutations in 42% (see **D**).

C. Clinical manifestations

Frequent acute bleeding episodes (**1**) resulting from minor trauma will result in severe functional consequences such as stiff knee joints (**2**) or elbows, or extensive soft-tissue hematoma. Severe manifestation with clotting factor less than 1% of normal occurs in 60%, moderate (1–5% F8 levels) in 15%, and mild (6–30% F8 levels) in 25% of patients. (Photographs obtained from Google, 2012.)

D. Factor VIII inversion

A frequent cause is an inversion between an unrelated gene A in intron 22 and one of two homologous distal genes A, followed by nonhomologous crossing-over. This flip tip inversion results in disruption of the *F8* gene between exon 22 and exon 23.

Further Reading

Brody H. Haemophilia. Nature 2014; 515–S157 (Nature Outlook: several articles)

Dahlbäck B. Blood coagulation. Lancet 2000; 355 (9215): 1627–1632

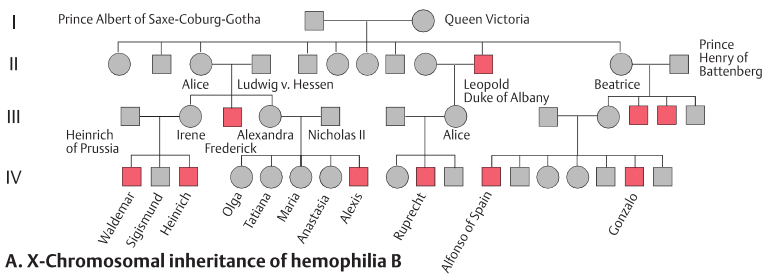
Gitschier J, et al. Detection and sequence of mutations in the factor VIII gene of haemophiliacs. Nature 1985; 315(6018): 427–430

Graw J, et al. Haemophilia A: from mutation analysis to new therapies. Nat Rev Genet 2005; 6(6): 488–501

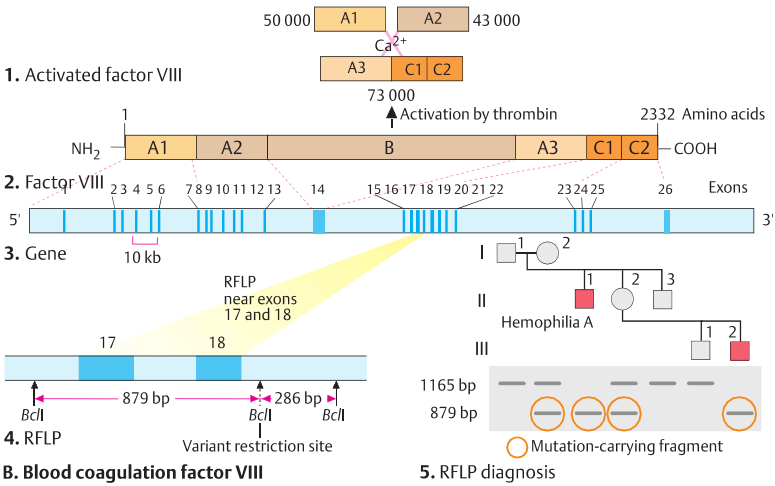
Hopff F. Über die Haemophilie oder die erbliche Anlage zu tötlichen Blutungen [inaugural dissertation]. Würzburg: Universität Würzburg, 1828

Pipe SW, ed. The Hemophilia Report. 2014. Available at: www.hemophiliareport.com

Rogaev EI, et al. Genotype analysis identifies the cause of the "royal disease." Science 2009; 326 (5954): 817



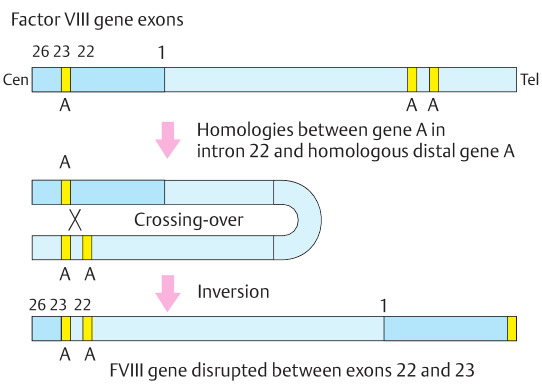
A. X-Chromosomal inheritance of hemophilia B



B. Blood coagulation factor VIII

1. Acute bleeding
2. Chronic consequences

C. Clinical manifestations



D. Factor VIII gene inversion

von Willebrand Bleeding Disease

von Willebrand disease (vWD, 193400) is the most common hereditary bleeding disorder. It results from dysfunction of a complex, multimeric glycoprotein (von Willebrand factor, vWF) in blood plasma, thrombocytes (platelets), and subendothelial mesenchymal tissue of blood vessels. It is a heterogeneous group of disorders, also referred to as von Willebrand–Jürgens syndrome. vWF has two basic biological functions: (1) it binds as a major adhesion molecule to specific receptors on the surface of platelets and subendothelial connective tissue by forming bridges between platelets and damaged regions of a vessel; (2) it binds to clotting factor VIII, and stabilizes it by prolonging its half-life in circulation. Deficiency of vWF reduces or abolishes platelet adhesion and leads to secondary deficiency of factor VIII. The frequency of hereditary deficiency of vWF is approximately 1:250 for all forms and approximately 1:8,000 for severe forms. It was first described in 1926 in a large family on the Åland islands in Finland by Erik von Willebrand.

A. von Willebrand factor glycoprotein

vWF is formed in endothelial cells, in megakaryocytes, and possibly in some other tissues. It is encoded by a large 178-kb gene with 52 exons of various sizes on chromosome 12p13.31 (613160). Along its 8.7-kb cDNA, several polymorphic restriction sites (red arrows) exist (**1**). From the mRNA, a primary peptide (prepro-vWF) of 2,813 amino acids is translated (**2**). Prepro-vWF is a polypeptide of 2,813 amino acids with a signal peptide of 22 amino acids and 5 repetitive functional domains, A to D and CK, which are distributed from the amino terminal in the sequence D1, D2, D3, A1, A2, A3, D4, B1, B2, B3, C1, C2, and CK. A segment of 741 amino acids in D1 and D2 corresponds to the von Willebrand antigen II. The various domains contain binding sites for factor VIII, heparin, collagen, thrombocytes, and thrombin. A tetrapeptide sequence Arg-Gly-Asp-Ser (RGDS) located near the C-terminus serves as a binding site for vWF. vWF contains 8.3% cysteine (234 of 2,813 amino acids), concentrated at the amino and carboxy ends, whereas the three A domains are cysteine-poor (**3**). After posttranslational modification, the mature plasma vWF contains 12 oligosaccharide side chains (**4**), resulting in a carbohydrate weight proportion of 19% of the vWF molecule.

B. Maturation of vWF

A prepropeptide is translated from the vWF mRNA. After the signal peptide is removed in the endoplasmic reticulum, the two pro-vWF units attach to each other at their carboxy ends by means of numerous disulfide bridges to form a dimer. The dimers represent the repetitive units of mature vWF. The pro-vWF dimers are transported to the Golgi apparatus, where the pro-vWF antigen (von Willebrand antigen II [vWAgII]) is removed.

C. Classification

vWD is classified into three subtypes. In types I and III, the defect is quantitative; in type II the defect is qualitative. Dominant and recessive (277480) phenotypes with vWF deficiency are often difficult to distinguish. Type I with subtypes A and B is the most frequent group (70% of all patients). The bleeding time is prolonged, but coagulation time is normal. Bleeding occurs mainly into mucocutaneous tissues rather than joints. (Figures adapted from Sadler, 2001.)

Further Reading

Israels SJ, et al. Inherited disorders of platelet function and challenges to diagnosis of mucocutaneous bleeding. Haemophilia 2010; 16(Suppl 5): 152–159

Konkle B. Disorders of platelets and vessel wall. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 965–973

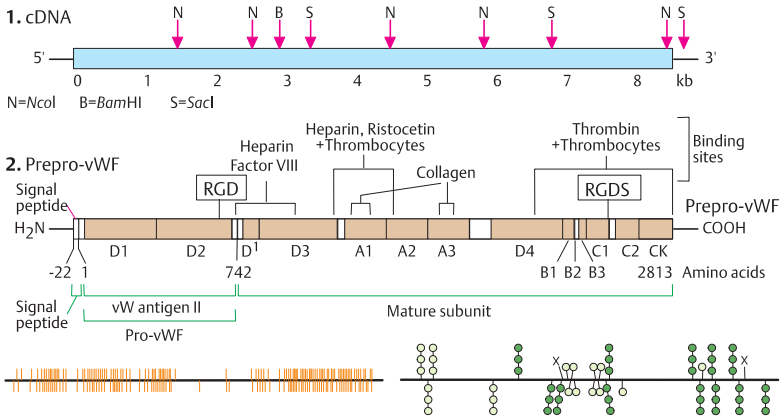
NIH. The diagnosis, evaluation and management of von Willebrand disease. Available at: www.nhlbi.nih.gov/health-pro/guidelines/current/von-willebrand-guidelines (accessed 16 August 2017)

Sadler JE, et al. Working Party on von Willebrand Disease Classification. Update on the pathophysiology and classification of von Willebrand disease: a report of the Subcommittee on von Willebrand Factor. J Thromb Haemost 2006; 4(10): 2103–2114

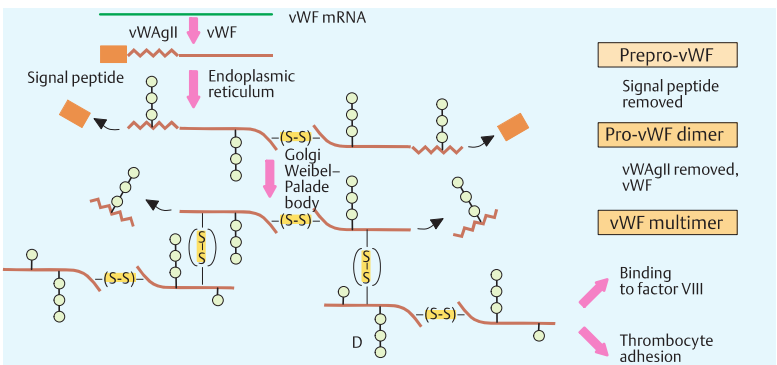
Sadler JE. Von Willebrand disease. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 4415–4431

von Willebrand EA. Hereditär pseudohemofili. Fin Lakaresallsk Handl 1926; 68: 87–112

von Willebrand EA, et al. Über ein neues vererbbares Blutungsübel. Dtsch Arch Klin Med 1933; 175: 453–483



A. von Willebrand cDNA and prepropeptide



B. Biosynthesis of von Willebrand factors (vWF)

von Willebrand disease	Genetics	vWF antigen	Factor VIII	Multimer structure
Type I	AD	Decreased	Decreased	Normal
IIA	AD	Decreased or normal	Decreased or normal	Large and intermediary absent
IIB	AD	Decreased or normal	Decreased or normal	Large absent in plasma, normal thrombocytes
IIC	AR	Decreased or normal	Decreased or normal	Large absent in plasma and in thrombocytes
IID	AD	Normal	Normal	Large absent
IIE	AD	Decreased	Normal	Large absent
III	AR	Absent	Greatly decreased	Absent

C. Classification of von Willebrand disease

Pharmacogenetics

Multiple individual genetic variants of the genome result in individual differences in pharmacological responses. This may result in either ineffective therapy or adverse reactions, even death. *Pharmacogenetics*, a term introduced by Motulsky (1957) and Vogel (1959), refers to a blend of genetics and pharmacology, now extended to *pharmacogenomics* to include the entire genome. Inherited variants of approximately 20 genes affect approximately 80 medications requiring attention in practice.

A. Malignant hyperthermia (MH)

Malignant hyperthermia (MH, 145600) is an autosomal dominant severe, life-threatening complication of anesthesia. It occurs in persons who are hypersensitive to halothane and similar agents used in general anesthesia. Normally, a nerve impulse depolarizes the plasma membrane of a nerve ending at the nerve–muscle end plate (**1**). The influx of calcium (Ca^{2+}) into the cell triggers the release of acetylcholine. This temporarily opens the receptor-controlled cation (sodium, Na^{+}) channels. This opens calcium channels located in the sarcoplasmic reticulum of the muscle cell and causes the myofibrils to contract. The calcium channels in the sarcoplasmic reticulum are regulated by the ryanodine receptor (**2**). Mutations in the ryanodine receptor, a protein with four transmembrane domains, cause greatly increased sensitivity to halothane and other anesthetic agents (**3**). They induce muscle rigidity, drastic elevation of temperature (hyperthermia), acidosis, and cardiac arrest (**4**). MH is inherited as an autosomal dominant trait (**5**). MH constitutes a group of at least seven genetically different disorders (MH1–MH7). The main gene involved encodes the skeletal muscle ryanodine receptor (RYR1, 180901) on chromosome 19q13.2. RYR2 (180902) and RYR3 (180903) encode the cardia and brain ryanodine receptors, respectively. Other genes responsible for MH are listed in OMIM 145600 under Allelic Series.

B. Butyrylcholinesterase deficiency

Prolonged apnea resulting from excessive muscle relaxation after administration of suxamethonium (succinylcholine) is the main manifestation of this adverse pharmacogenetic reaction. It affects approximately 1 in 200 individuals, and 1 in 100,000 are affected by the severe form. Affected individuals do not degrade suxamethonium because of insufficient activity of serum butyrylcholinesterase (formerly acetylcholinesterase 177400). This enzyme hydrolyzes butyrylcholine more readily than acetylcholine (OMIM 100740). Individuals at risk cannot be identified by enzyme activity alone (**1**). However, all three genotypes can be determined after dibucaine, an inhibitory substance, is added (**2**). Individuals at risk (red squares) have only 20% activity of butyrylcholinesterase, heterozygotes have 50 to 70%, and homozygous normal persons approximately 80%. Butyrylcholinesterase is encoded by the *BCHE* gene located at 3q26.1. (Figure adapted from Harris, 1975.)

C. Examples of genetically determined adverse reactions to pharmaceuticals

Seven examples of the numerous human pharmacogenetic disorders are listed.

Further Reading

Evans WE, et al. Pharmacogenomics--drug disposition, drug targets, and side effects. N Engl J Med 2003; 348(6): 538–549

Harris H. The Principles of Biochemical Genetics. 2nd ed. Amsterdam: Holland Publishing Company, 1975

Hughes D, et al. Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. Nat Rev Genet 2015; 16(8): 459–471

Motulsky AG. Drug reactions enzymes, and biochemical genetics. J Am Med Assoc 1957; 165(7): 835–837

Nebert DW. Pharmacogenetics and pharmacogenomics. Nature Encyclopedia Human Genome 2003; 4: 558–567

Nebert DW, Vesell ES. Pharmacogenetics and Pharmacogenomics. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medial Genetics. 6th ed. Philadelphia, PA: Churchill Livingstone-Elsevier, 2013

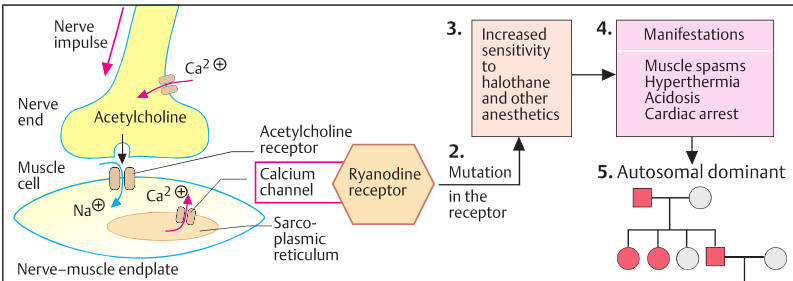
Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature 2015; 526(7573): 343–350

Robinson R, et al. Mutations in RYR1 in malignant hyperthermia and central core disease. Hum Mutat 2006; 27(10): 977–989

Tyner JW. Integrating functional genomics to accelerate mechanistic personalized medicine. Cold Spring Harb Mol Case Stud 2017; 3(2): a001370

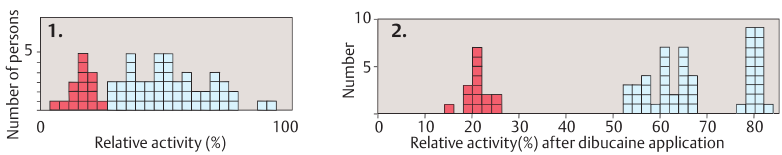
Vogel F. Moderne Probleme der Humangenetik. Ergeb Inn Med Kinderheilkd 1959; 12: 52–125

Weinshilboum R. Inheritance and drug response. N Engl J Med 2003; 348(6): 529–537



1. Activated calcium channels at the activated nerve–muscle (motor) endplate

A. Malignant hyperthermia due to a calcium channel disorder in muscle cells



B. Butyrylcholinesterase deficiency

Defect	Relevant chemical substance	Clinical consequence	Frequency	Pathogenesis	Genetics
Coumarin resistance	Coumarin (warfarin)	Ineffective anticoagulation therapy	Rarer than 1:80 000	Increased vitamin K affinity due to enzyme or receptor defect	Autosomal dominant
Increased sensitivity to isoniazid	Isoniazid Sulfamethazine Phenelzine Hydralazine, etc.	Polyneuritis, lupus-like reaction	In about 50%	Decreased activity of liver isoniazid acetylase	Autosomal recessive
Isoniazid ineffective	Isoniazid Sulfamethazine Phenelzine Hydralazine	Reduced antituberculous effect		Increased isoniazid excretion	Autosomal dominant
Glucose-6-phosphate dehydrogenase (G6PD) deficiency	Sulfonamides Antimalarial drugs Nitrofurantoin Vicia faba	Hemolysis	Rare in Europeans, frequent in Africa and parts of SE Asia	G6PD deficiency in erythrocytes	X-chromosomal (many mutant forms)
Hemoglobin Zürich	Sulfonamides	Hemolysis	Rare	Unstable hemoglobin due to point mutation in β-globin (arginine instead of histidine in position 63)	Autosomal dominant
Hemoglobin H	Sulfonamides	Hemolysis	Rare	Unstable hemoglobin of 4 β chains due to deletion of the α loci	Autosomal dominant
Glaucoma in adults (some forms)	Corticoids	Glaucoma	Frequent	Unknown	Possibly autosomal dominant

C. Examples of genetically determined adverse reactions to pharmaceuticals

Cytochrome P450 (CYP) Genes

The cytochrome P450 system refers to a large gene family (*CYP* genes; OMIM 108330). *CYP* genes encode enzymes with different functions in detoxification of various chemical substances. They derive their name from their maximal light absorption at 450 nm after binding to carbon monoxide. Cytochrome P450 enzymes form a large, evolutionarily related family of proteins with different enzymatic specificity in mammals. They degrade complex chemical substances, such as drugs or plant toxins, by an oxidation system (monooxygenases) in the microsomes of the liver and mitochondria of the adrenal cortex.

A. Cytochrome P450 system

The cytochrome P450 enzymes carry out phase I of a detoxification pathway (**1**): a substrate (RH) is oxidized to ROH utilizing atmospheric oxygen (O_2), with water (H_2O) formed as a byproduct. A reductase delivers hydrogen ions (H^+) from either NADPH or NADH. In phase II, ROH is further degraded and eliminated. The P450 enzymes have a wide spectrum of activity (**2**). Characteristically, a single P450 protein can oxidize several structurally different chemical substances, or several P450 enzymes can degrade a single chemical substrate. The enzyme activities of phases I and II have to be well coordinated, since toxic intermediates occasionally arise in the initial stages of phase II.

B. Debrisoquine metabolism

Debrisoquine is an isoquinoline-carboxamidine. It was used to treat high blood pressure until it was found to cause severe side effects in 5 to 10% of the population (OMIM 608902). Affected individuals have reduced activity of debrisoquine-4-hydroxylase (*CYP2D6*). This enzyme degrades several pharmacological substances such as β-adrenergic blockers, antiarrhythmics, and antidepressives. Two groups can be distinguished in the population: those with normal and those with slow degradation (**1**). Individuals with low activity are at increased risk of untoward toxic reactions. Individuals with a slow rate of degradation have an increased ratio of debrisoquine/4-hydrodebrisoquine. This enzyme is encoded by the *CYP2D6* gene (OMIM 124030) located at 22q13.1. Certain mutations of the primary transcript of this gene, with nine exons, produce aberrant splicing (**2**). As a result, variant mRNAs contain an intron and produce proteins with reduced enzymatic activity. (Figure adapted from Gonzalez et al., 1988.)

C. CYP gene superfamily

The cytochrome P450 (CYP) genes in mammals consist of a superfamily of genes that resemble each other in exon/intron structure and that encode related enzymes. The CYP gene family arose during the last 1,500 to 2,000 million years. The largest P450 family in mammals is CYP2, with 16 genes in humans. It is assumed that the CYP2 family developed in response to toxic substances in plants. At least 30 gene duplications and gene conversions have led to an unusually diverse repertoire of CYP genes. Important enzymes for drug metabolism are CYP2C8, CYP2C9, CYP2C18, and CYP2C19, which together metabolize more than 50 compounds. CYP3A4, CYP2D6, and CYP2C9 are responsible for 50, 25, and 5% of drug metabolism, respectively. (Figure adapted from Gonzalez et al., 1988, and Gonzalez and Nebert, 1990.)

Further Reading

Genetics Home Reference. CYP gene family. Available at: http://ghr.nlm.nih.gov/geneFamily/cyp. Accessed March 1, 2017

Gonzalez FJ, et al. Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. Nature 1988; 331(6155): 442–446

Gonzalez FJ, Nebert DW. Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. Trends Genet 1990; 6(6): 182–186

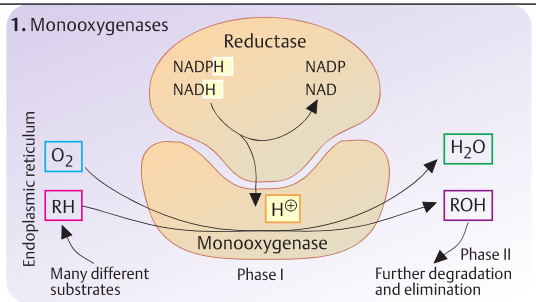
Lynch T, Price A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. Am Fam Physician 2007; 76(3): 391–396

Nebert DW, et al. Clinical importance of the cytochromes P450. Lancet 2002; 360(9340): 1155–1162

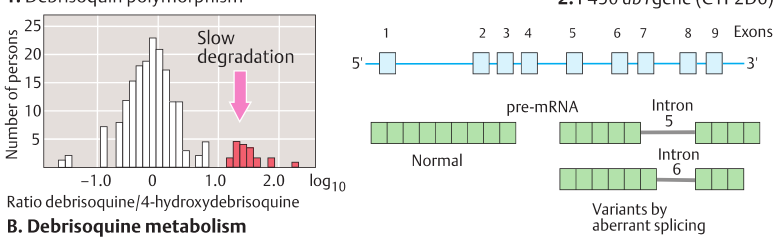
Nebert DW, et al. Cytochrome P450 *(CYP) gene* superfamily. Nature Encyc Hum Genome 2003; 1: 1028–1037

Nelson DR, et al. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. Pharmacogenetics 2004; 14(1): 1–18

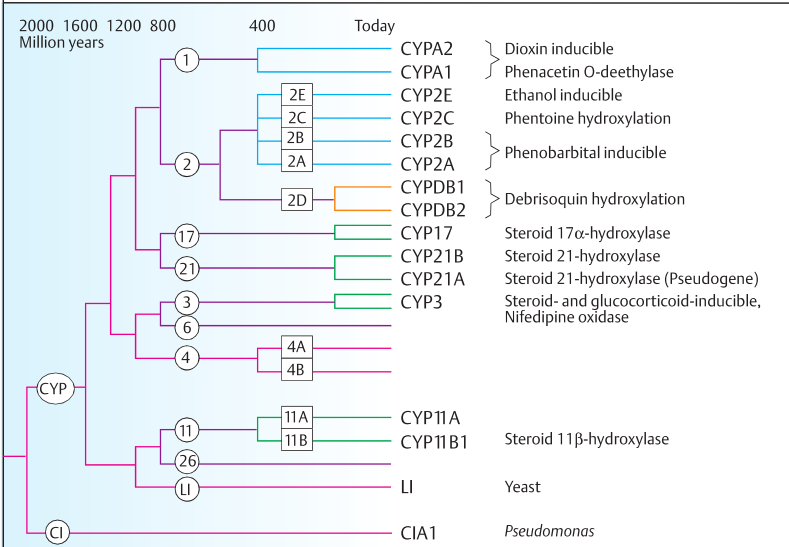
Nelson DR, et al. The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. Philos Trans R Soc Lond B Biol Sci 2013; 368(1612): 20120474



A. Cytochrome P450 system



B. Debrisoquin metabolism



C. CYP gene superfamily

Genetics of Diabetes Mellitus

Diabetes mellitus (DM) is an etiologically heterogeneous group of genetic subtypes with a wide spectrum of phenotypic (clinical) manifestations. DM is defined by elevated fasting levels of glucose in the blood (>125 mg/dL, hyperglycemia). Two principal types of DM are insulin-dependent type 1 (1TDM, 222100) and non-insulin-dependent type 2 (2TDM, 125853). T1DM results from autoimmune or infectious processes destroying the beta cells in the pancreas. T2DM has multiple causes. It is a common health problem in many parts of the world, with up to 10% of the population being at risk or affected.

A. Insulin biosynthesis

Human insulin is encoded by a gene (176730) with a leading signal sequence (L) and two coding exons in the 5′ direction of the gene, located on the short arm of chromosome 11 (11p15.5). It has a beta-cell-specific enhancer and a variable number of tandem repeats (VNTRs) upstream (5′) of the gene. The primary transcript is spliced into mRNA and translated into preproinsulin (1,430 amino acids). After the signal sequence (24 amino acids) and the C-peptide are removed, biosynthesis is completed. The gene is expressed exclusively in the beta cells of the pancreas.

B. Insulin receptor

Insulin exerts its multiple functions through its receptor (INSR 147670). INSR is a tetramer of two extracellular α subunits and two transmembrane β subunits, joined by disulfide bridges at specific sites. Distinct functional domains in both extracellular and intracellular parts of the receptor reflect the various functions. Insulin receptor substrate IRS and Shc proteins mediate the principal functions of insulin: activating growth factors, protein synthesis, glycogen synthesis, and glucose transport. The *INSR* gene located at 19p13.2 spans more than 120 kb and has 22 exons, 11 for each of the α and β chains.

C. Etiology of diabetes mellitus

T1DM is mainly caused by external factors, such as certain viral infections, which directly or by autoimmune reactions cause destruction of beta cells on a background of individual genetic susceptibility. T2DM has a strong genetic background. Fuchsberger et al. (2016) identified 18 coding variants at 13 loci to exceed genome-wide significance (GWAS, see p. 220). All were attributed to common variants rather than rare variants. T2DM is influenced by obesity as a strong environmental component, including influences in early life. Monozygotic twins with T1DM are concordant in approximately 25%, and for T2DM in approximately 40 to 50%. Several types of autosomal dominant forms of T2DM exist in adults (maturity onset diabetes of the young, MODY types 1–6; 125850). DM is a secondary manifestation of several genetically determined diseases involving insulin receptor defects (insulin resistance syndromes, see OMIM 125853).

D. Genetic susceptibility loci

A major influence on the genetic susceptibility to T1DM is exerted by certain alleles of class I MHC genes (see p. 306). Several alleles of the DR3 and DR4 loci are associated with susceptibility to T1DM, especially DR3/DR4 heterozygotes. Some alleles of DR2 confer relative resistance to DM. Genetic and genome-wide association studies have identified 36 regions or loci in the human genome conferring susceptibility to T2DM type 2 (listed in OMIM 125853).

Further Reading

Daneman D. Type 1 diabetes. Lancet 2006; 367 (9513): 847–858

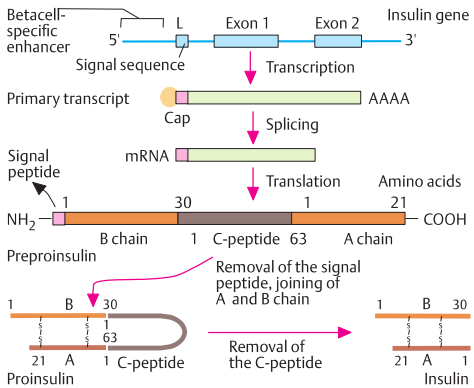
Diabetes Consortium. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014; 46(3): 234–244

Brody H. Diabetes Outlook. Nature 2012; 485: S1–S19

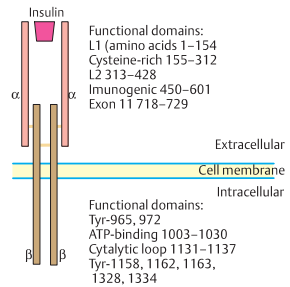
Franks PW, McCarrthy MI. Exposing the exposures responsible for type 2 diabetes and obesity. Science 2016; 354(6308): 69–73

Fuchsberger C, et al. The genetic architecture of type 2 diabetes. Nature 2016; 536(7614): 41–47

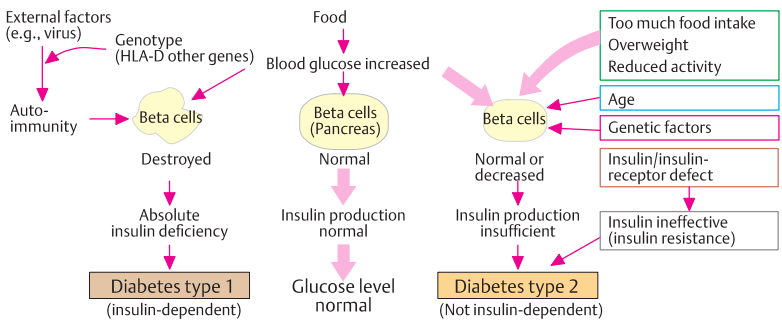
Stumvoll M, et al. Type 2 diabetes: principles of pathogenesis and therapy. Lancet 2005; 365 (9467): 1333–1346



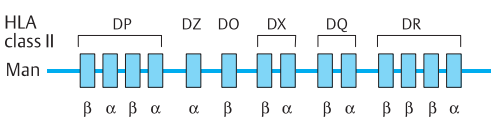
A. Insulin biosynthesis



B. Insulin receptor



C. Etiology of Diabetes mellitus



Susceptible haplotypes: *DR3* and/or *DR4 DQA1*0331, DQB1*0302, DQA1*501, DQB1*0201*
(present in 40% of children with type 1A diabetes mellitus vs. 2% in the population)

Protective haplotypes: *DQA1*0102, DQB1*0602*

Other susceptibility loci: Xp11.23-q13.3, 12q24.2, 1p13, 6p21.3

1. Insulin-dependent diabetes mellitus type 1 (OMIM 222100)

Susceptibility loci:
2q24.1
2q32
5q34-q35.2
6p12
6q22-q23
11p12-p11.2
12q24.2
13q12.1
13q34
17cen-q21.3
17q25
19p13.2
19q13.1-q13.2
20q12-q13.1 and others

1. Non-insulin-dependent diabetes mellitus type 2 (OMIM 125853)

D. Genetic susceptibility loci

Amino Acid Degradation and Urea Cycle Disorders

Mutations in genes encoding the various enzymes involved in amino acid metabolism and urea cycle cause a large group of disorders. Two examples are presented here.

A. Phenylalanine degrading system

Phenylketonuria (PKU, 261600) is an autosomal recessive deficiency of the enzyme phenylalanine hydroxylase (PAH). This enzyme hydroxylates phenylalanine (Phe) to tyrosine. PKU results from mutations in the *PAH* gene.
The PAH gene (OMIM 612349), located on chromosome 12q23.2, has 13 exons spanning 90 kb of DNA, with complex 5′ untranslated *cis*-acting, *trans*-activated regulatory elements. The gene shows developmental and tissue-specific transcription and translation. The hepatic and renal gene product is a 452-amino-acid polypeptide. Several polymorphic sites (RFLPs and SNPs, and a tetranucleotide short tandem repeat) are in linkage disequilibrium and define haplotypes. The PAH protein has catalytic, regulatory, and tetramerization domains. Modifying genes influence the phenotype. Hyperphenylalaninemia is defined as a plasma Phe concentration of >120 µM (2 mg/dL). Long-term exposure to Phe concentrations >600 µM results in severe mental retardation, as in classical PKU. This condition, first described in 1934 by Asbjørn Følling in Norway, is the prototype of a treatable metabolic disease. Here, dietary phenylalanine restriction must be instituted in the neonatal period. PKU is detected by neonatal screening. Maternal PKU causes various developmental problems if the maternal Phe level is not well controlled throughout pregnancy.
The complex Phe hydroxylating system includes tetrahydrobiopterin (BH_4) cofactor, which requires several enzymes for recycling, including dihydropteridine reductase (261630) and pterin-4α-carbinolamine dehydratase (264070).

B. Distribution of PAH mutations

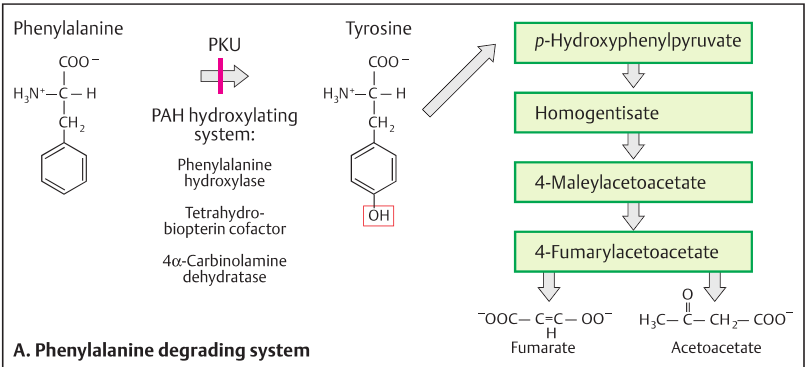
More than 500 disease-causing mutations have been recorded in all 13 exons and flanking sequences. Many vary in frequency in different populations. PKU is rare in Finnish, Ashkenazi, American aboriginal, and Japanese populations. (Data from Scriver, 2007, and Zschocke, 2003.)

C. Urea cycle defects

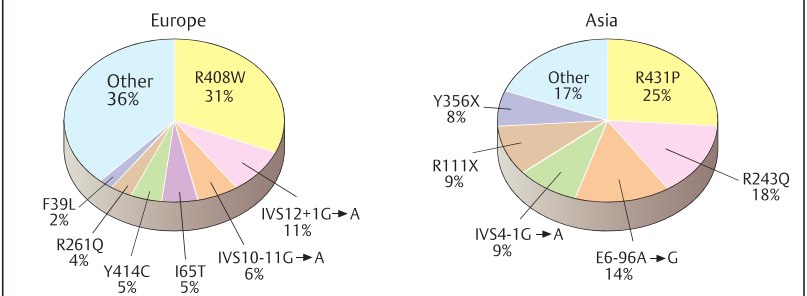
Terrestrial vertebrates synthesize urea by means of the urea cycle, the first metabolic pathway to be described (by Krebs and Henseleit, a medical student, in 1932). Five metabolic disorders characterized by toxic high plasma levels of ammonium result from defects in the genes encoding enzymes regulating the urea cycle (indicated by red bars in the figure part **C**). Arginine, the immediate precursor of urea, is hydrolyzed by arginase (207830) to urea and ornithine. Ornithine transcarbamoylase (OTC, 310461) transfers carbamoyl phosphate to ornithine, which results in citrulline. Carbamoyl phosphate is synthesized from NH_{4+}, CO_2, H_2O, and ATP by carbamoyl phosphate synthetase (237300). Argininosuccinate synthetase (215700) catalyzes the condensation of citrulline and aspartate to argininosuccinate. This is cleaved by argininosuccinase (202900) into arginine and fumarate. Affected children may show progressive lethargy and coma leading to death in the neonatal period. The most common form is X-linked OTC resulting from mutations in the OTC gene located at Xp21.1.

Further Reading

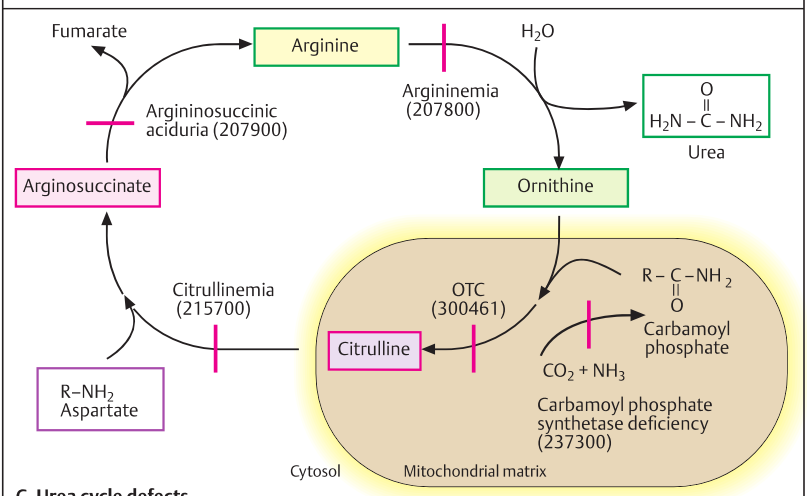
Ah Mew N, et al. Urea Cycle Disorders Consortium of the Rare Diseases Clinical Research Network. Clinical outcomes of neonatal onset proximal versus distal urea cycle disorders do not differ. J Pediatr 2013; 162(2): 324–9.e1
Ah Mew N, et al. Urea cycle disorder overview. In: Pagon RA, et al., eds. GeneReviews. Available at: https://www.ncbi.nlm.nih.gov/books/NBK1217/. Accessed March 4, 2017
Blau N, et al. Phenylketonuria. Lancet 2010; 376 (9750): 1417–1427
Brusilow SW, et al. Urea cycle enzymes. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 1909–1963
Mitchell JJ, et al. Phenylalanine hydroxylase deficiency. Genet Med 2011; 13(8): 697–707
PAH Database. Available at: www.pahdb.mcgill.ca/. Accessed March 4, 2017
Scriver CR. The PAH gene, phenylketonuria, and a paradigm shift. Hum Mutat 2007; 28(9): 831–845
Scriver CR, et al. Hyperphenylalaninemia: phenylalanine hydroxylase deficiency. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. Available at: www.ommbid.com/
Zschocke J. Phenylketonuria mutations in Europe. Hum Mutat 2003; 21(4): 345–356



A. Phenylalanine degrading system



B. Distribution of PAH mutations in different populations



C. Urea cycle defects

Cholesterol Biosynthesis Pathway

Several hereditary diseases result from mutations in genes encoding enzymes of the cholesterol biosynthesis pathway. Cholesterol is a precursor of many steroid hormones and a major constituent modulating the fluidity of cell membranes in eukaryotes. In 1932, Wieland and Dane elucidated its structure as a monosaturated 27-carbon sterol. The biosynthetic pathway of cholesterol requires approximately 30 enzymatic reactions regulated by 22 genes in a series including oxidation with molecular oxygen, reductions, demethylations, and alterations in double bonds. Konrad Bloch was awarded the Nobel Prize in 1954 for elucidating this pathway.

A. Malformation syndromes due to defects in cholesterol metabolism

Approximately six different genetic diseases are known to result from a block of the cholesterol biosynthesis pathway (see next page). Three examples are shown: (**1**) the autosomal recessive Smith–Lemli–Opitz syndrome (270400); (**2**) X-linked chondrodysplasia punctata type 2 (CDPX2, Conradi–Hünermann syndrome, 302960); and (**3**) autosomal dominant Greenberg skeletal dysplasia (215140). For more details see table 10 in the appendix (p. 410). (Illustrations in 1, kindly provided by the parents of the child; 2, courtesy of Dr Richard I. Kelley, Baltimore, Maryland, United States; 3, courtesy of the late Dr David L. Rimoin, Los Angeles, California, United States.)

B. Cholesterol biosynthesis overview

Cholesterol biosynthesis begins with acetyl coenzyme A (acetyl-CoA), from which all 27 carbon atoms are derived. Acetyl-CoA and acetoacetyl-CoA condense to 3-hydroxy-3-methylglutaryl-CoA. This is converted by 3-hydroxy-3-methylglutaryl-CoA reductase to mevalonate. This is the precursor of isoprene, which is synthesized in three steps (not shown). Squalene, a 30-carbon linear isoprenoid, is synthesized from six isoprene units. Isopentyl pyrophosphate is the starting point of a reaction C5 → C10 → C15 → C30. The distal (postsqualene) part of the cholesterol biosynthesis pathway begins with squalene.

Mevalonic aciduria (610377) results from a block in mevalonate kinase (251170). This variable autosomal recessive disease is characterized by increased urinary excretion of mevalonic acid associated with failure to thrive, psychomotor retardation, vomiting, diarrhea, episodes of fever, and dysmorphic facial features.

C. Squalene to lanosterol

Initially, squalene is circularized through a reactive intermediate, squalene epoxide (not shown), to lanosterol, the first postsqualene sterol intermediate. Squalene epoxide is closed by a cyclase to lanosterol, a 30-carbon sterol. This requires movements of electrons through four double bonds and the migration of two methyl groups. Removal of the 24–25 double bond results in dihydrolanosterol, which is the other precursor of cholesterol.

Further Reading

Fitzky BU, et al. Mutations in the Delta7-sterol reductase gene in patients with the Smith-Lemli-Opitz syndrome. Proc Natl Acad Sci U S A 1998; 95(14): 8181–8186

Goldstein JL, Brown MS. Regulation of the mevalonate pathway. Nature 1990; 343(6257): 425–430

Greenberg CR, et al. A new autosomal recessive lethal chondrodystrophy with congenital hydrops. Am J Med Genet 1988; 29(3): 623–632

Herman GE. Disorders of cholesterol biosynthesis: prototypic metabolic malformation syndromes. Hum Mol Genet 2003; 12(Spec No 1): R75–R88

Kelley RI, et al. Abnormal sterol metabolism in patients with Conradi-Hünermann-Happle syndrome and sporadic lethal chondrodysplasia punctata. Am J Med Genet 1999; 83(3): 213–219

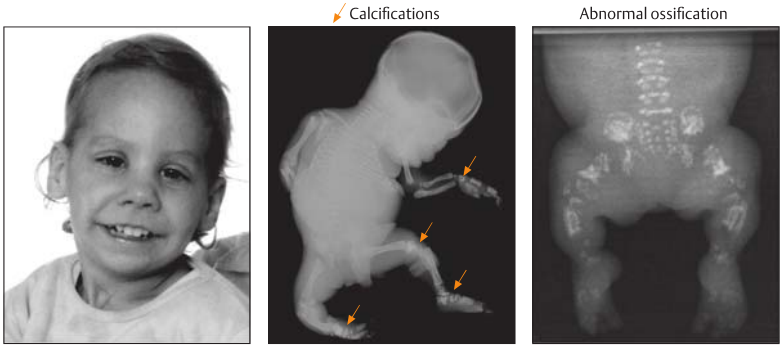
Koo G, et al. Discordant phenotype and sterol biochemistry in Smith-Lemli-Opitz syndrome. Am J Med Genet A 2010; 152A(8): 2094–2098

Smith DW, Lemli L, Opitz JM. A newly recognized syndrome of multiple congenital anomalies. J Pediatr 1964; 64: 210–217

Waterham HR. Inherited disorders of cholesterol biosynthesis. Clin Genet 2002; 61(6): 393–403

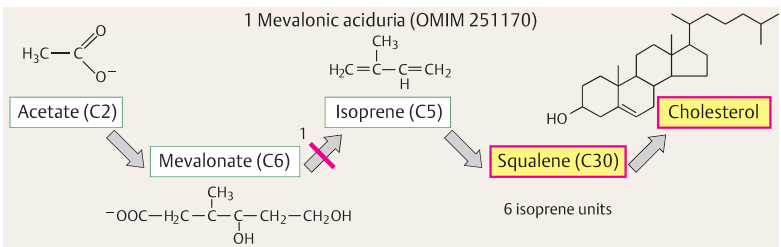
Witsch-Baumgartner M, et al. Age and origin of major Smith-Lemli-Opitz syndrome (SLOS) mutations in European populations. J Med Genet 2008; 45(4): 200–209

Witsch-Baumgartner M, et al. Maternal apo E genotype is a modifier of the Smith-Lemli-Opitz syndrome. J Med Genet 2004; 41(8): 577–584

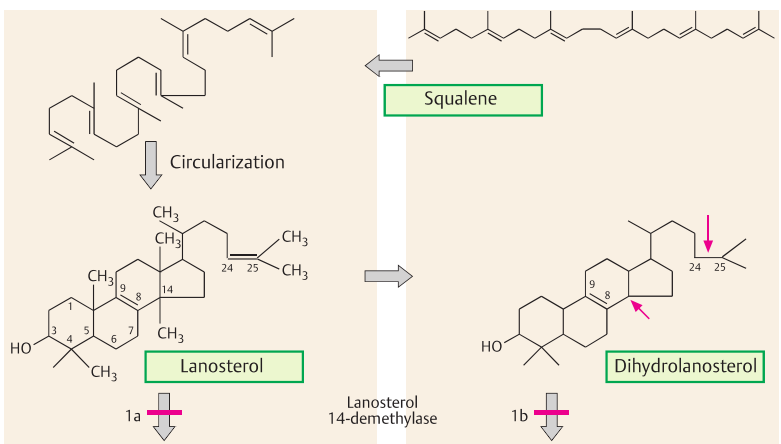


1. Smith–Lemli–Opitz syndrome **2.** Chondrodysplasia punctata **3.** Greenberg dysplasia

A. Malformation syndromes due to defects in cholesterol metabolism (examples)



B. Cholesterol biosynthesis (overview)



C. Squalene to lanosterol, the first step in the distal pathway

Distal Cholesterol Biosynthesis Pathway

In the distal part of the cholesterol biosynthesis pathway (postsqualene), lanosterol and dihydrolanosterol (see previous page) are converted to desmosterol and 7-dihydrocholesterol (7-DHC), the immediate precursors of cholesterol. Mutations in genes encoding the enzymes required for cholesterol biosynthesis have been found at all steps. They cause rare genetic diseases characterized by developmental delay and abnormalities of the skeletal and other systems.

A. Distal cholesterol biosynthesis pathway and diseases

Lanosterol and dihydrolanosterol (see previous page) are converted by four enzymatic reactions through four intermediate metabolites to the immediate precursors of cholesterol, desmosterol (cholesta-5(6),24-dien-3β-ol), and 7-dihydrocholesterol (7-DHC). The enzymatic steps remove three methyl groups at C-4 and C-14, open one double bond at C-24, and shift the C-8–C-9 double bond to C-7–C-8 by an isomerase. Some of the enzymatic reactions must occur in a defined sequence: the Δ^8-Δ^7 isomerization must follow the C-14α demethylation. The pathway is tied to different cellular functions and signaling pathways. Lanosterol and its two subsequent intermediates, 4,4-dimethyl-cholesta-8(9),14,24-trien-3β-ol and 4,4-dimethyl-cholesta-8(9),24-dien-3 β-ol, have meiosis-stimulating activity and accumulate in the ovary and testis (see Herman, 2003). 7-DHC is the direct precursor of vitamin D. Hedgehog signaling proteins are modified by cholesterol (see p. 192).

Please note that Antley-Bixler syndrome with genital anomalies and disordered steroidogenesis (OMIM 201750, gene located at 7q11.23) and without genital anomalies or disordered steroidogenesis (OMIM 207410, gene FGFR2 located at 10q26.13) is heterogeneous. It is not a disorder of primary cholesterol biosynthesis.

Seven genetic diseases caused by defects in the distal (postsqualene) cholesterol biosynthesis pathway are known. These are (in descending order of the pathway reactions): (1) a proportion of patients with Antley–Bixler syndrome (201750), (2) the prenatally lethal Greenberg skeletal dysplasia (215140), (3) X-chromosomal CHILD syndrome (congenital hemidysplasia with ichthyosiform erythroderma or nevus and limb defects, 308050), (4) X-chromosomal dominant chondrodysplasia punctata type 2 (Conradi–Hünermann syndrome, 302960), (5) lathosterolosis (607330), (6) Smith–Lemli–Opitz syndrome (270400), and (7) desmosterolosis (602398). The main features of these diseases are summarized in table 10 in the appendix (p. 410).

Further Reading

Farese RV Jr, et al. Cholesterol metabolism and embryogenesis. Trends Genet 1998; 14(3): 115–120

Greenberg CR, et al. A new autosomal recessive lethal chondrodystrophy with congenital hydrops. Am J Med Genet 1988; 29(3): 623–632

Herman GE. Disorders of cholesterol biosynthesis: prototypic metabolic malformation syndromes. Hum Mol Genet 2003; 12(Spec No 1): R75–R88

Herman GE, et al. Characterization of mutations in 22 females with X-linked dominant chondrodysplasia punctata (Happle syndrome). Genet Med 2002; 4 (6): 434–438

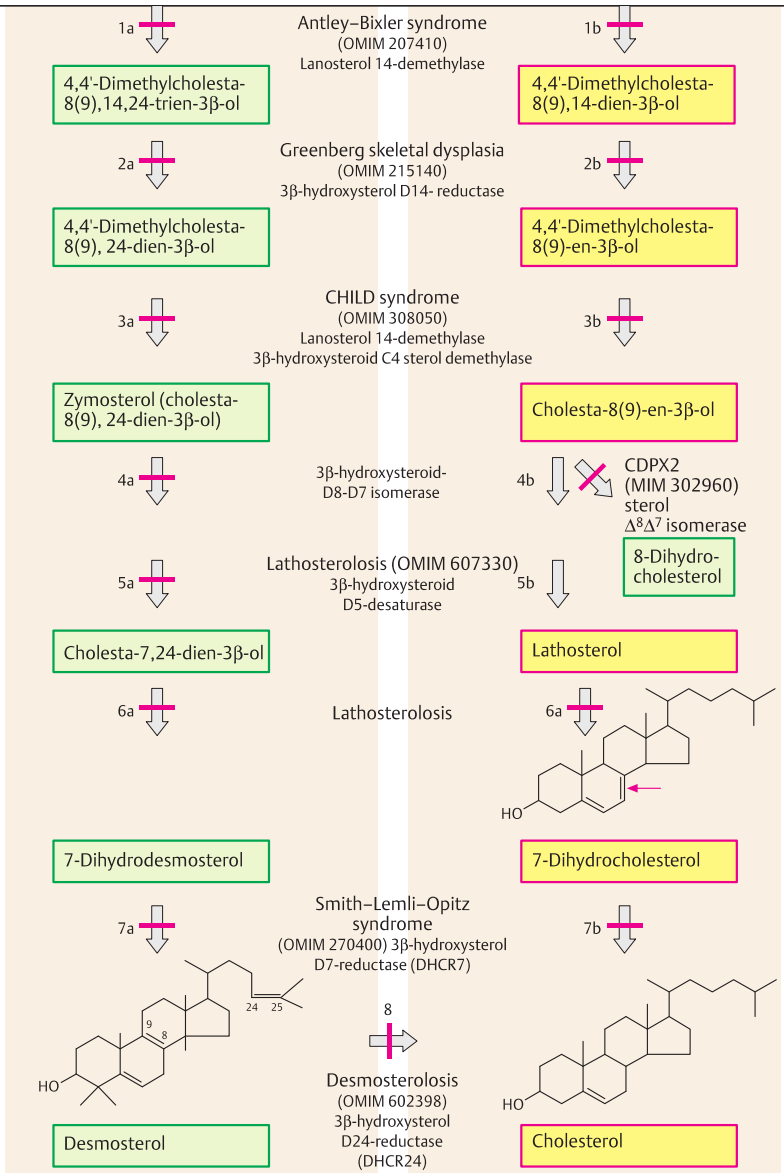
Kelley RI, Herman GE. Inborn errors of sterol biosynthesis. Annu Rev Genomics Hum Genet 2001; 2: 299–341

Kelley RI, Hennekam RCM. Smith–Lemli–Opitz syndrome. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 6183–6201

Schaaf CP, et al. Desmosterolosis-phenotypic and molecular characterization of a third case and review of the literature. Am J Med Genet A 2011; 155A(7): 1597–1604

Waterham HR. Inherited disorders of cholesterol biosynthesis. Clin Genet 2002; 61(6): 393–403

Waterham HR, et al. Autosomal recessive HEM/Greenberg skeletal dysplasia is caused by 3 beta-hydroxysterol delta 14-reductase deficiency due to mutations in the lamin B receptor gene. Am J Hum Genet 2003; 72(4): 1013–1017



A. Distal cholesterol biosynthesis pathway and diseases

Familial Hypercholesterolemia

Familial hypercholesterolemia (FH, 143890) is an autosomal dominant disorder caused by heterozygous mutations in the gene encoding the low-density lipoprotein receptor (*LDLR*, 606945) located on 19p13.2. FH is characterized by elevated plasma levels of cholesterol bound to low-density lipoprotein (LDL), the main cholesterol-transporting lipoprotein. Triglyceride levels are normal. Lipoproteins are classified according to increasing density: chylomicrons, very-low-density lipoproteins (VLDLs), low-density lipoproteins (LDLs), and high-density lipoproteins (HDLs). A related form is hypercholesterolemia type B (144010) due to mutations in the apolipoprotein B gene (107730).

A. The disease phenotype

The general features of FH are listed in part 1. It occurs, in the heterozygous state, in approximately 1 in 500 individuals (**1**). Plasma cholesterol levels are greatly increased (**2**). The activity of functional LDL receptors per cell is decreased by approximately 50% and is nearly absent in homozygotes (**3**). Important clinical signs are deposits of cholesterol esters in the tendons, especially the Achilles tendon, and the skin (xanthomas, **4**) and lipid deposits around the iris in the eye (**5**). In the rare homozygous state (1 in 10^6), FH leads to death during the first or second decade of life. Several other related forms at other gene loci are known (OMIM 143890). (Figure in part 2 and 3 adapted from Goldstein et al., 2001; 4 and 5 are the author's observations.)

B. The LDL receptor

Heterozygous mutations in the LDL receptor gene cause FH (see next page). The gene consists of 18 exons spanning 45 kb of genomic DNA. It is transcribed into 5.3-kb mRNA. The LDL receptor is a membrane-bound 160-kDa protein cell surface receptor of 839 amino acids with 6 distinct functional domains. Seven cysteine-rich units of 40 amino acids each form the ligand-binding region (exons 2–6). The epidermal growth factor (EGF) precursor homology domain (exons 7–14) allows the lipoprotein to dissociate from the receptor in the endosome (see next page). The intracellular domain encoded by part of exon 17 and the 5′ end of exon 18 contains signals for localizing the receptor in the coated pits during endocytosis and targeting within hepatocytes. (Figure adapted from Goldstein et al., 2001, according to Hobbs et al., 1990.)

C. LDL receptor–mediated endocytosis

The LDL receptor mediates the endocytosis of LDL. Receptors loaded with LDL accumulate in a coated pit (**1**), which forms an endocytotic vesicle inside the cell (**2**). (Photographs from Anderson et al., 1977.)

D. Homology with other proteins

The mammalian LDL receptor is highly conserved in evolution (90% identical within mammals, 79% identity between man and shark), dating back at least 500 million years. The proximal halves of the extracellular domains of the LDL receptor family are structurally related to the epidermal growth factor (EGF) family (p. 284). The families of genes encoding the LDL receptor and EGF are related to proteins of the blood coagulation system: factors IX (134540) and X (613872), protein C (612283), and complement C9 (120940). Other related genes (not shown) are those encoding the VLDL (very-low-density lipoprotein) receptor, the APOE receptor 2 (ApoER2, the LDL receptor–related protein (LRP1, 107770), and LDL receptor protein 2 (LRP2, megalin, 600073).

Further Reading

Anderson RGW, et al. Role of the coated endocytic vesicle in the uptake of receptor-bound low density lipoprotein in human fibroblasts. Cell 1977; 10 (3): 351–364

Brown MS, et al. A receptor-mediated pathway for cholesterol homeostasis. Science 1986; 232 (4746): 34–47

Goldstein JL, et al. Familial hypercholesterolemia. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 2863–2913

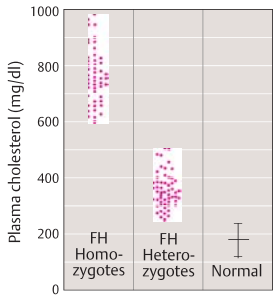
Hobbs HH et al. The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. Ann Rev Genet 1990; 24: 133–170

Low Density Lipoprotein Familial Hypercholesterolemia Database, University College London. Available at: http://www.ucl.ac.uk/ldlr. Accessed March 6, 2017

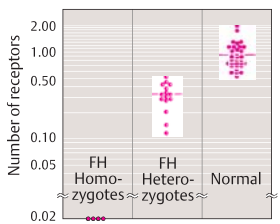
Rader DJ, et al. Disorders of lipoprotein metabolism. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 3145–3161

– Low-density lipoprotein (LDL) and cholesterol elevated in blood plasma
– Premature arteriosclerosis
– Xanthoma in skin and tendons
– Decreased life expectancy
– Autosomal dominant
– Mutation in LDL receptor gene

1. General features



2. Hypercholesterolemia

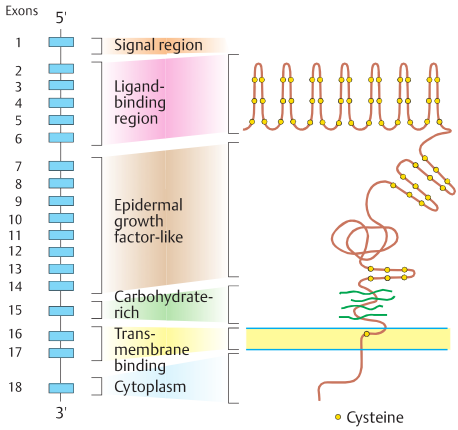


3. LDL receptors decreased

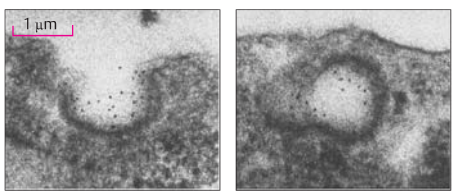
4. Xanthoma formation

5. Arcus lipoides

A. Familial hypercholesterolemia



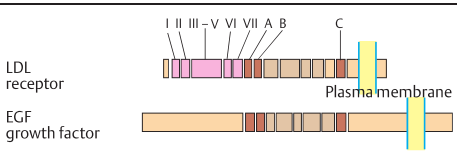
B. LDL receptor

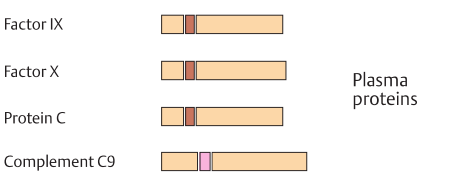


1. Coated pit **2.** Endocytotic vesicle

Electron micrographs of fibroblasts in culture that have taken up LDL molecules (black dots, made visible by binding to ferritin).

C. Receptor-mediated endocytosis of LDL





D. Homology with other proteins

LDL Receptor Mutations

Mutations in the low-density lipoprotein (LDL) receptor occur throughout the gene. LDL is the main transport agent of cholesterol in the blood. Its hydrophobic core contains approximately 1,500 esterified cholesterol molecules surrounded by an outer layer of phospholipids and unesterified cholesterols containing a single apoB-100 lipoprotein molecule. LDL delivers cholesterol to peripheral tissues and regulates de novo cholesterol synthesis in these tissues.

A. LDL receptor classes of mutation

The effects of LDL receptor mutations can be understood on the basis of LDL receptor metabolism. The receptor–LDL complex enters the cell by endocytosis (see next page) within a coated vesicle. The receptor is separated from the LDL including the apoB-100 in the endosome. The LDL receptor is recycled to the cell surface. In the lysosome, the LDL is broken down into amino acids and cholesterol. Free cholesterol activates the enzyme acetyl-CoA cholesterol transferase (ACAT), which catalyzes the esterification. The key enzyme for endogenous cholesterol synthesis is 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase). This enzyme is downregulated by exogenous LDL uptake. LDL receptor mutations interrupt this feedback mechanism and result in increased endogenous cholesterol synthesis. Five principal classes of LDL receptor mutations can be distinguished: (**1**) receptor null mutations (R^0) resulting from lack of receptor protein synthesis in the endoplasmic reticulum, (**2**) defective intracellular transport to the Golgi apparatus, (**3**) defective extracellular ligand binding, (**4**) defective endocytosis (R^+ mutations), and (**5**) failure to release the LDL molecules inside the endosome (recycling-defective mutations). (Figure adapted from Goldstein et al., 2001.)

B. Mutational spectrum

A 5′ signal sequence and five functional domains can be distinguished along the LDLR gene (606945) as shown (EGF, epidermal growth factor). More than 1,700 unique mutations have been recorded in the LDL receptor gene. Of these, 65% are missense mutations caused by substitutions, 24% are small DNA rearrangements, and 11% are large rearrangements. Mutations occur in all parts of the gene, but there is a relative excess of mutations in exons 4 and 9. A high proportion of mutations (74%) located in the ligand-binding domain (exons 2–6) involve amino acids conserved in evolution. Alu repeats may be involved as a cause of intragenic deletions. Some mutations differ in frequency in different populations. (Figure adapted from Goldstein et al., 2001.)

C. A mutation in the LDL receptor gene

This example of partial analysis demonstrates a mutation in exon 9. First, exon 9 was amplified by PCR (P1, primer 1; P2 primer 2). The mutation in codon 408, GTG (valine) to GTA (methionine), produces a recognition site (N) for *Nla*III (GATC) that is not normally present. This results in two fragments of 126 and 96 bp instead of the usual 222-bp fragment (**1**). Thus, affected individuals (1 and 3 in the pedigree) have two smaller fragments of 126 and 96 kb in addition to the 222-kb fragment (**2**). Partial DNA sequence analysis of the patient (individual 1 in the pedigree) demonstrates the mutation by the presence of an additional adenine (A, yellow circle) next to the normal guanine (**3**). On this basis, other members of a family can easily be examined for presence or absence of the mutation. Today, the entire gene may be sequenced. (Photographs kindly provided by Dr H. Schuster, Berlin.)

Further Reading

Brown MS, et al. Familial hypercholesterolemia. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Diseases. 2011. Available at: www.ommbid.com/OMMBID/

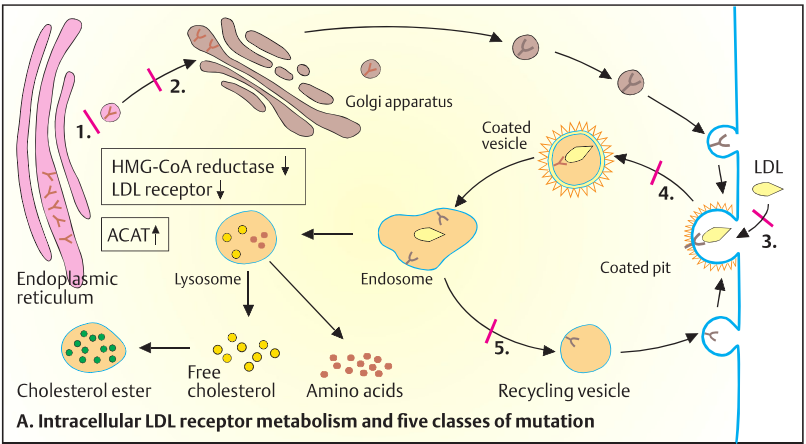
Familial Hypercholesterolemia (FH) Variant Databases. Available at: www.ucl.ac.uk/ugi/fh

Goldstein JL, Brown MS, Hobbs HH. Familial hypercholesterolemia. In: Scriver CR, et al, eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York: McGraw-Hill; 2001:2863–2913. (www.ommbid.com/)

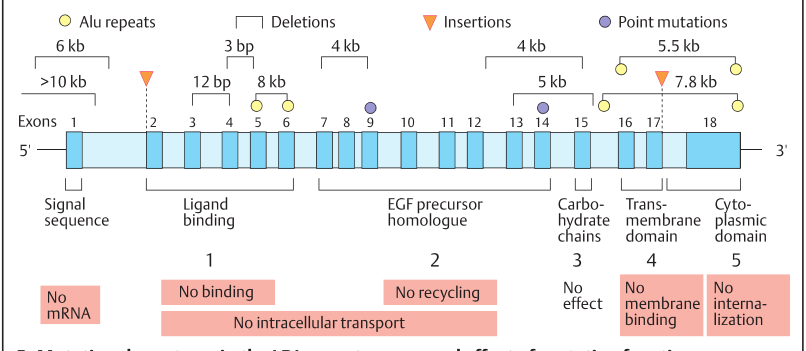
Hegele RA. Lipoprotein and lipid metabolism. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. Elsevier

Leigh SE, et al. Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. Ann Hum Genet 2008; 72(Pt 4): 485–498

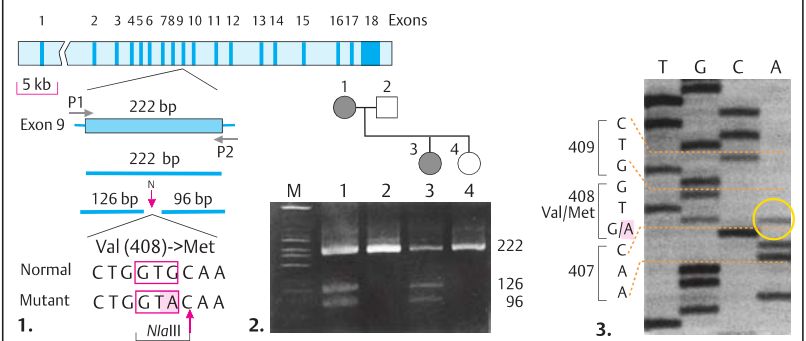
Rader DJ, et al. Disorders of lipoprotein metabolism. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 3145–3161



A. Intracellular LDL receptor metabolism and five classes of mutation



B. Mutational spectrum in the LDL receptor gene and effect of mutation function



C. Point mutation in the LDL receptor gene

Lysosomal Storage Disorders

Lysosomal disorders are a group of approximately 50 genetic diseases involving various functions of lysosomes. Lysosomes are membrane-enclosed intracellular vesicles with a diameter of 0.05 to 0.5 µm. They are required for the intracellular degradation of large molecules. They contain more than 50 active hydrolytic enzymes (acid hydrolases), such as glycosidases, sulfatases, phosphatases, lipases, phospholipases, proteases, and nucleases (collectively called lysosomal enzymes) in an acid milieu (approximately pH = 5). Lysosomal enzymes enter a lysosome by means of a recognition signal (mannose 6-phosphate) and a corresponding receptor. Lysosomal disorders are characterized by abnormal storage of different macromolecules. They are grouped according to the main class of stored material: glycogen, mucopolysaccharides, glycoproteins, glycolipids, sphingomyelin, gangliosides, and others. Different forms of therapy using enzyme replacement are being developed for some of lysosomal disorders.

A. Receptor-mediated endocytosis and lysosome formation

Extracellular macromolecules to be degraded are taken into the cell by endocytosis. First, the molecules are bound to specific cell surface receptors (receptor-mediated endocytosis). The loaded receptors are concentrated in an invagination of the plasma membrane (coated pit). This separates from the plasma membrane and forms a membrane-enclosed cytoplasmic compartment (coated vesicle). The cytoplasmic lining of the vesicle consists of a network of a trimeric protein, clathrin. The clathrin coat is removed within the cell, forming an endosome. The receptor and the molecule to be degraded (the ligand) are separated and the receptor is recycled to the cell surface.

A multivesicular body (endolysosome) forms and takes up acid hydrolases arriving in clathrin-enclosed vesicles. Hydrolytic degradation takes place in the lysosome. Parts of the membrane are also recycled.

B. Mannose 6-phosphate receptors

A mannose 6-phosphate receptor serves as a recognition signal for uptake into the endolysosome, which will also be recycled back into the Golgi apparatus. The acid milieu in the lysosomes is maintained by a hydrogen pump in the membrane that hydrolyzes ATP and uses the energy produced to move hydrogen ions into the lysosome. Some of the mannose 6-phosphate (mannose 6-P) receptors are transported back to the Golgi apparatus.

Two types of mannose 6-phosphate receptor molecules exist. They differ in their binding properties and their cation dependence. They consist of either 2 (cation-independent mannose 6-phosphate receptor, CI-MDR) or 16 (cation-dependent mannose 6-phosphate receptor, CD-MDR) extracellular domains, with different numbers of amino acids. The cDNA of CI-MPR is identical to insulinlike growth factor 2 (IGF-2). Thus, CI-MPR is a multifunctional binding protein.

C. Biosynthesis

Two enzymes are essential for the biosynthesis of mannose 6-phosphate recognition signals: a phosphate transferase and a phosphoglycosidase. The phosphate is delivered by uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) to uridine-5′-diphosphate-N-acetylglucosamine-glycoprotein-*N*-acetylglucosaminyl-phospho-transferase (GlcNAc-phospho-trans-ferase). A second enzyme (*N*-acetylglucosamine-1-phosphodiester-*N*-acetyl-glucosaminidase, GlcNAc-phosphoglycosidase) removes the *N*-acetylglucosamine, leaving the phosphate residue at position 6 of the mannose. (Figures adapted from de Duve, 1984, and Sabatini and Adesnik, 2001.)

Further Reading

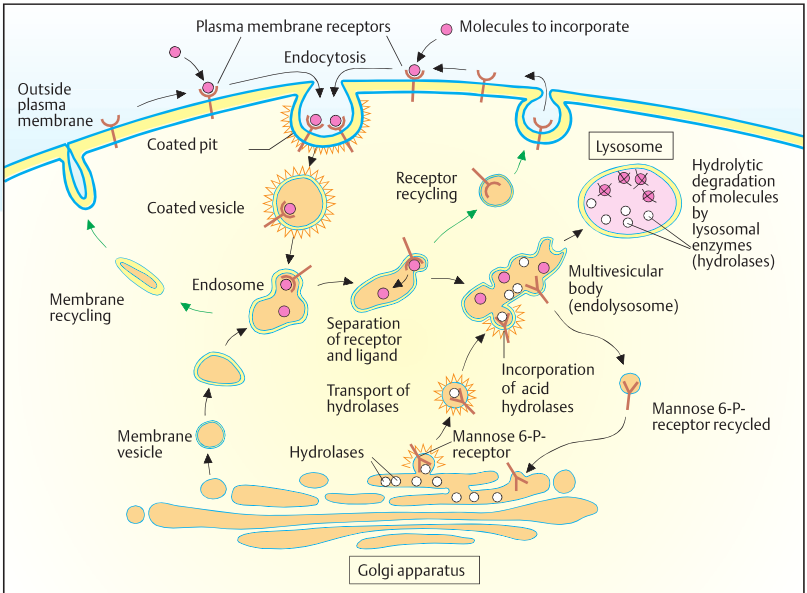
de Duve C. A Guided Tour of the Living Cell. Vols. 1 and 2. New York, NY: Scientific American Books, 1984

Fuller M et al. Epidemiology of lysosomal storage diseases. Chapter 2 in: Fabry Disease. Mehta A et al, eds. Oxford: Oxford PharmaGenesis, 2006

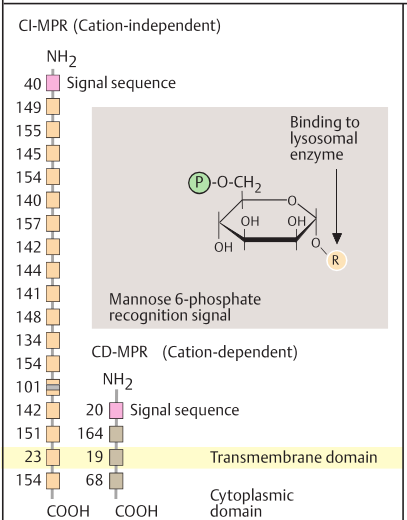
Hopkin RJ, et al. Lysosomal storage diseases. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 3191–3197

Kingma SD, et al. Epidemiology and diagnosis of lysosomal storage disorders; challenges of screening. Best Pract Res Clin Endocrinol Metab 2015; 29(2): 145–157

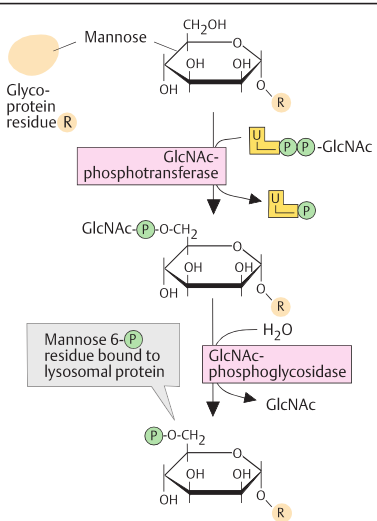
Sabatini DD, Adesnik MB. The biogenesis of membranes and organelles. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 433–517



A. Receptor-mediated endocytosis and lysosome biogenesis



B. Mannose 6-phosphate receptors (MPR)



C. Biosynthesis of mannose 6-phosphate recognition signal

Lysosomal Enzyme Defects

Genetic defects in enzymes degrading different macromolecules in lysosomes cause a large group of different diseases. Their clinical signs and biochemical and cellular manifestations depend on the function normally performed by the enzyme involved. Macromolecules that are usually degraded remain in lysosomes and accumulate in the cell, resulting in lysosomal storage diseases. This occurs at different rates, so that each disease has its own characteristic course. Twelve groups of genetically determined disorders of specific lysosomal functions are known, each with approximately 3 to 10 individually defined diseases.

A. Defective uptake of enzymes into lysosomes: I-cell disease

Mucolipidosis type II alpha/beta (252500), also called *I-cell disease* because of conspicuous cytoplasmic inclusions, first described by Leroy and DeMars in 1967, is a severe, progressive autosomal recessive disorder of abnormal lysosomal transport and protein sorting in mesenchymal cells. The first step of a two-step reaction in the Golgi apparatus is defective because of homozygous or compound heterozygous mutations in the *GNPTAB* gene (607840), located at 12q23.2, encoding the lysosomal enzyme N-acetylglucosamine-1-phosphotransferase, α and β subunits (GlcNAc-phosphotransferase, see previous page). This enzyme catalyzes the first step in the synthesis of the mannose 6-phosphate determinant for targeting hydrolases into the lysosome. As a result, the recognition marker that binds mannose 6-phosphate is lacking and mucolipids accumulate in mesenchymal cells (**1**), but not in normal fibroblasts (**2**). The vesicular inclusions consist of hydrolases that cannot enter the lysosomes because the mannose 6-phosphate recognition signal is absent. Lysosomes lack several enzymes, whereas the concentration of these enzymes outside the cells is increased. Mucolipidoses usually become apparent in the first 6 months of life (**3**). Two complementation groups have been delineated. An allelic disorder, mucolipidosis III alpha/beta is also caused by mutation in the *GNPTAB* gene (252600), and mucolipidosis III gamma is caused by mutations in the *GNPTG* gene (252605).

B. Degradation of heparan sulfate

Lysosomal enzymes are bond specific, not substrate specific. Thus, they also degrade other glycosaminoglycans, such as dermatan sulfate, keratan sulfate, and chondroitin sulfate (mucopolysaccharides). Ten specific enzyme defects cause the mucopolysaccharide storage diseases (see next page). Heparan sulfate is an example of a macromolecule that is degraded stepwise by eight different lysosomal enzymes. The first step in heparan sulfate degradation is the removal of sulfate from the terminal iduronate group by an iduronate 2-sulfatase. A defect in the gene encoding this enzyme leads to the X-chromosomal mucopolysaccharide storage disease type II (MPS II, type Hunter). The second enzymatic step removes the terminal iduronate by an α-L-iduronidase. If this is defective, MPS I results. Enzyme defects in the next five steps cause genetically different forms of autosomal recessive MPS IIIA, IIIC, IIIB, and IIID as shown (see next page). MPS type VII is caused by a defect in β-glucuronidase (step 7). It differs in phenotype from MPS types I, II, and III.

Further Reading

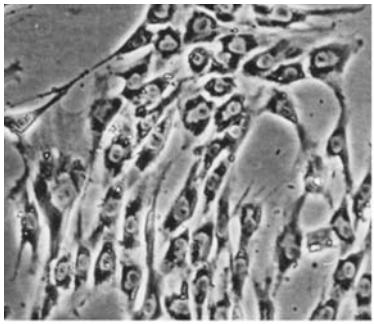
Cathey SS, et al. Molecular order in mucolipidosis II and III nomenclature. Am J Med Genet A 2008; 146A(4): 512–513

Kornfeld S, Sly WS. I-cell disease and Pseudo-Hurler polydystrophy: disorders of lysosomal enzyme phosphorylation and localization. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 3469–3482

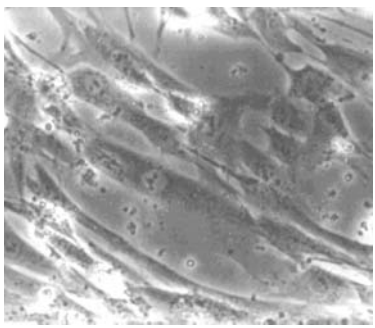
Leroy JG, Demars RI. Mutant enzymatic and cytological phenotypes in cultured human fibroblasts. Science 1967; 157(3790): 804–806

Marschner K, et al. A key enzyme in the biogenesis of lysosomes is a protease that regulates cholesterol metabolism. Science 2011; 333(6038): 87–90

Tiede S, et al. Mucolipidosis II is caused by mutations in GNPTA encoding the alpha/beta GlcNAc-1-phosphotransferase. Nat Med 2005; 11(10): 1109–1112



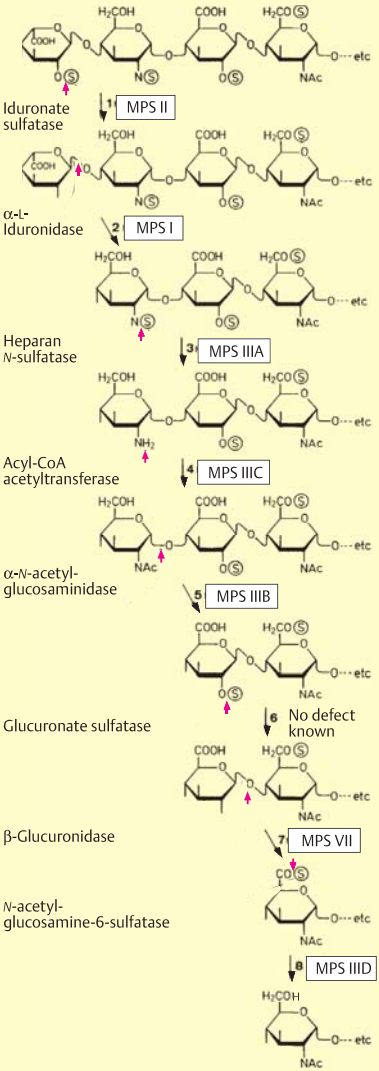
1. Fibroblast culture in I-cell disease



2. Normal fibroblast culture

3. Patient with I-cell disease

A. Defective uptake of enzymes in lysosomes: I-cell disease



B. Degradation of heparan sulfate by eight lysosomal enzymes

Mucopolysaccharide Storage Diseases

The mucopolysaccharide storage diseases (mucopolysaccharidoses) are a clinically and genetically heterogeneous group of 10 different genetic types (see phenotypic series in OMIM 253200) caused by defects in different enzymes for mucopolysaccharide degradation (glycosaminoglycans). All defects are transmitted by autosomal recessive inheritance, except for mucopolysaccharide storage disease type II (Hunter; OMIM 309900) (see table 11 in the appendix, p. 411).

A. Mucopolysaccharide storage disease type I (Hurler)

Young infants with mucopolysaccharide storage disease type I (Hurler, MPS IH) (252800) seem normal at first. Early signs of the disease occur at approximately 1 to 2 years of age, with increasing coarsening of the facial features, retarded mental development, limited joint mobility, enlarged liver, umbilical hernia, and other signs. Radiographs show coarsening of skeletal structures (dysostosis multiplex). The photographs show the same patient at different ages (author's photographs). MPS IS (Scheie) is a clinically different, less severe allelic disease.

B. Mucopolysaccharide storage disease type II (Hunter)

This type of mucopolysaccharidosis is transmitted by X-chromosomal inheritance (309900). Four cousins from one pedigree are shown in the diagram. Clinically, the disease is similar to, but less rapidly progressive than, MPS I. (Photographs from Passarge et al., 1974.)

Diagnosis and therapy

The diagnosis of MPS is based on the patient's history, clinical and radiological evaluation, and increased urinary concentration of one of several types of glycosaminoglycan, depending on the type of MPS. Molecular genetic diagnosis is possible in most forms. Different forms of enzyme substitution exist.

Further Reading

Hopkin RJ, et al. Lysosomal storage diseases. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 3191–3197

Kakkis ED. Enzyme replacement therapy for the mucopolysaccharide storage disorders. Expert Opin Investig Drugs 2002; 11(5): 675–685

Neufeld EF, et al. The mucopolysaccharidoses. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 3421–3452

OMIM (Online Mendelian Inheritance of Man). Available at: www.ncbi.nlm.nih.gov/omim

Passarge E, et al. Diseases caused by genetic defects in lysosomal muco-polysaccharide-catabolism. Mucopolysaccharidoses [in German]. Dtsch Med Wochenschr 1974; 99(4): 144–155

Ponder KP, et al. Gene therapy for mucopolysaccharidosis. Expert Opin Biol Ther 2007; 7(9): 1333–1345

Tolar J, et al. Combination of enzyme replacement and hematopoietic stem cell transplantation as therapy for Hurler syndrome. Bone Marrow Transplant 2008; 41(6): 531–535

Wolf DA, et al. Gene therapy for neurologic manifestations of mucopolysaccharidoses. Expert Opin Drug Deliv 2015; 12(2): 283–296

Wraith JE. Mucopolysaccharidoses. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medial Genetics. 6th ed. Philadelphia, PA: Churchill Livingstone-Elsevier, 2013

8 weeks

7 months

$2\frac{1}{4}$ years

$3\frac{3}{4}$ years

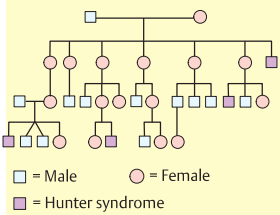
5 years

8 years

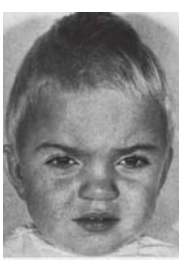
Dysostosis multiplex

Joint contractures

A. Mucopolysaccharide storage disease type I (Hurler)



X-Chromosomal inheritance


$4\frac{1}{2}$ years

10 years

13 years

21 years

B. Mucopolysaccharide storage disease type II (Hunter)

Peroxisomal Disorders

Peroxisomal disorders are a group of 20 mainly autosomal recessive disorders due to defective enzyme or a transporter protein in peroxisomes or dysfunction of component in peroxisomal biogenesis (OMIM 170993). Major clinical features are neonatal hypotonicity, craniofacial abnormalities, failure to thrive, and other features. Peroxisomes are small membrane-bound intracellular organelles of approximately 0.5- to 1.0-µm diameter, somewhat smaller than mitochondria. Their name is derived from hydrogen peroxide, which is formed as an intermediary product of oxidative metabolism. Most cells, especially in the liver and kidney, contain approximately 100 to 1,000 peroxisomes. A peroxisome is surrounded by a single-layer granular matrix, which contains approximately 50 to 100 different matrix enzymes. These are involved in anabolic and catabolic metabolic functions, such as β-oxidation of fatty acids, biosynthesis of phospholipids and bile acids, and others. Peroxisome biogenesis involves the synthesis of matrix proteins (peroxins, PEX1, 3,5,7,10, 11A, 11B, 11G, 12, 13, 14, 16, 19, 26 see OMIM 602136). and their receptor-mediated transfer into the organelle under the control of *PEX* genes and peroxisomal targeting signals (PTS).

A. Biochemical reactions

The electron micrograph (**1**) shows three peroxisomes in a rat liver cell. The dark striated structures within the organelles are urates, a result of an enzyme that oxidizes uric acid. Peroxisomes have both catabolic (degrading) and anabolic (synthesizing) functions (**2**). Two biochemical reactions are especially important: a peroxisomal respiratory chain and the β-oxidation of very-long-chain fatty acids. In the peroxisomal respiratory chain (**3**), certain oxidases and catalases act together. Specific substrates of the oxidases are organic metabolites of intermediary metabolism. Very-long-chain fatty acids are broken down by β-oxidation (**4**) in a cycle with four enzymatic reactions. Energy production in peroxisomes is relatively inefficient compared with that of mitochondria. While free energy in mitochondria is mainly preserved in the form of ATP (adenosine triphosphate), in peroxisomes it is mostly converted into heat. Peroxisomes are probably a very early adaption of living organisms to oxygen. (Photograph from de Duve, 1984.)

B. Peroxisomal diseases

Six important examples of autosomal recessive peroxisomal diseases are listed, together with OMIM numbers. Patients with neonatal adrenoleukodystrophy do not form sufficient amounts of plasmalogens and cannot adequately degrade phytanic acid and pipecolic acid.

C. Zellweger cerebrohepatorenal syndrome

This is a group of 13 recessive diseases resulting from mutations in *PEX* genes (see phenotypic series in OMIM 214100). It is recognized by a characteristic facial appearance (**1–4**), extreme muscle weakness (**5**), and several accompanying manifestations such as calcified stippling of the joints on radiographs (**6**), renal cysts (**7, 8**), and clouding of the lens and cornea. The severe form of the disease usually leads to death before the age of 1 year. (Photographs 1–5 from Passarge and McAdams, 1967.)

Further Reading

Crane DI et al. PEX1 mutations in the Zellweger spectrum of the peroxisome biogeneisis disorders. Hum Mutat 2005; 26: 167–175

de Duve C. A Guided Tour through the Living Cell. New York, NY: Scientific American Books, 1984

Gould SJ, et al. The peroxisome biogenesis disorders. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 3181–3217

Muntau AC, et al. Defective peroxisome membrane synthesis due to mutations in human PEX3 causes Zellweger syndrome, complementation group G. Am J Hum Genet 2000; 67(4): 967–975

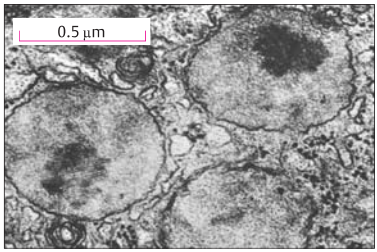
Passarge E, McAdams AJ. Cerebro-hepato-renal syndrome. A newly recognized hereditary disorder of multiple congenital defects, including sudanophilic leukodystrophy, cirrhosis of the liver, and polycystic kidneys. J Pediatr 1967; 71(5): 691–702

Peroxisome Database. Available at: http://www.peroxisomedb.org/. Accessed March 2017

Shimozawa N, et al. A human gene responsible for Zellweger syndrome that affects peroxisome assembly. Science 1992; 255(5048): 1132–1134

Wanders RJA. Peroxisomal disorders. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. Philadelphia, PA: Churchill Livingstone-Elsevier, 2013

Wanders RJA, Waterham HR. Biochemistry of mammalian peroxisomes revisited. Annu Rev Biochem 2006; 75: 295–332



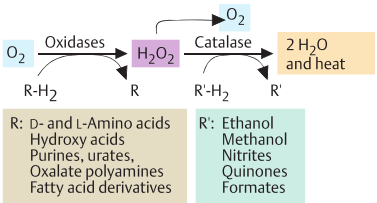
1. Peroxisomes in a rat liver cell

a) Catabolic
- H_2O_2-involving cellular respiration
- β-Oxidation of long-chain fatty acid,
- Prostaglandins, cholesterol side chains and others
- Purines, urates
- Pipecolic acid, dicarboxy acids
- Ethanol, methanol

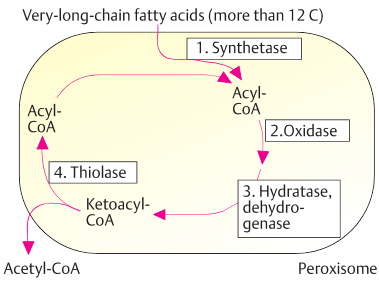
b) Anabolic
- Phospholipids (plasmalogen)
- Cholesterol, bile acids
- Gluconeogenesis
- Glyoxalate transamination

2. Function of peroxisomes

A. Biochemical reactions in peroxisomes



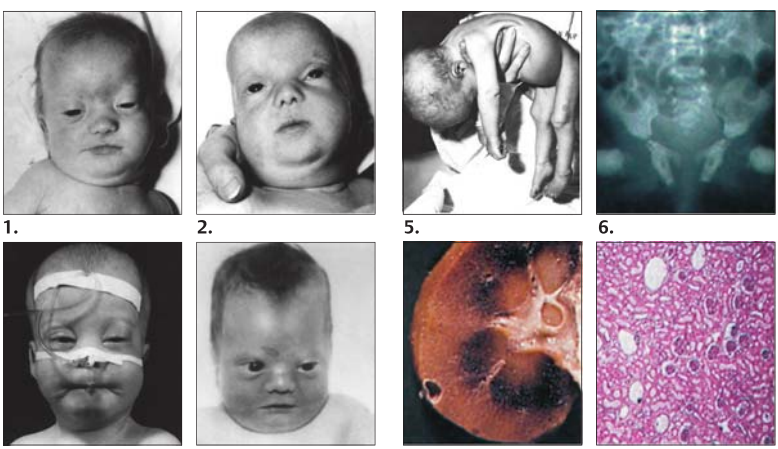
3. Peroxisomal respiratory chain



4. β-Oxidation

214100 Zellweger cerebrohepatorenal syndrome
202370 Neonatal adrenoleukodystrophy
266510 Infantile Refsum disease
239400 Hyperpipecolic acidemia
215100 Rhizomelic chondrodysplasia punctata
259900 Primary hyperoxaluria type I
and others

B. Examples of peroxisomal diseases



1. **2.** **5.** **6.**

3. **4.** **7.** **8.**

C. Zellweger cerebrohepatorenal syndrome

Components of the Immune System

The immune system consists of an adaptive and an innate system. The adaptive immune system is a relatively young evolutionary immune response mediated by T and by B lymphocytes. Three key properties of the adaptive immune system, present in vertebrates only, are: (1) a highly diverse repertoire of antigen receptors, (2) a memory function for antigens previously encountered, and (3) immune tolerance to avoid unwarranted damage to the host's own cells. The innate immune system, originated from invertebrates, consists of germline-encoded proteins defending against pathogen-associated molecular patterns. It reacts nonspecifically when encountering a pathogen. Recently, it was shown that the innate immune system also can build an immunological memory.

Three principal responses of the immune system to invading foreign molecules are: (1) neutralizing extracellular pathogens by antibodies, (2) destroying a cell that is infected, and (3) killing bacteria directly by macrophages.

A. Lymphatic organs

The primary lymphoid tissues are the thymus and bone marrow. Secondary lymphoid tissues are the lymph nodes, the spleen, and accessory lymphoid tissue (ALT) including tonsils and appendix.

B. Lymphocytes

The approximately 2×10^{12} lymphocytes in the human body are equal in mass to the brain or the liver. Their specific role in adaptive immunity was shown in the late 1950s by irradiation experiments. A mouse irradiated above a certain dose was no longer able to mount an immune response. However, the adaptive immune response could be restored by lymphocytes from a normal control mouse.

C. T cells and B cells

Two functionally different types of lymphocyte exist: T lymphocytes and B lymphocytes. Immature T lymphocytes differentiate in the thymus during embryonic and fetal development (thus designated T cells). B lymphocytes differentiate in the bone marrow in mammals and in the bursa of Fabricius in birds (thus designated B cells). Further maturation and differentiation take place in the lymph nodes (T cells) and in the spleen (B cells).

D. Cellular and humoral immune response

The first phase of the immune response to an antigen (e.g., a bacterium, virus, fungus, or foreign protein) is rapid proliferation of B cells (humoral immune response). Mature B cells develop into plasma cells, which secrete effector molecules, the antibodies (immunoglobulins). These interact with the antigen by binding to it.

E. Immunoglobulin molecules

The basic structural motif of an antibody molecule (immunoglobulin, Ig) is a **Y**-shaped protein composed of different polypeptide chains. A common type of Ig has two heavy chains (H chains) and two light chains (L chains). Both chains contain regions with variable sequences and constant sequences of amino acids. At defined sites, the chains are held together by disulfide bonds.

F. Antigen–antibody binding

An antibody molecule is bivalent by having two identical antigen-binding sites at the end of each arm of the **Y**-shaped end. The structure that an antibody recognizes is an antigenic determinant or epitope. Here a foreign molecule, the antigen, is recognized and firmly bound to six hypervariable regions (three from the light chain and three from the heavy chain). In the hypervariable regions, the amino acid sequences differ from one molecule to the next. As a result, the antibodies can bind a wide spectrum of different antigenic molecules.

(Figures adapted from Alberts et al., 2015.)

Further Reading

Abbas AK, et al. Cellular and Molecular Immunology. 8th ed. Philadelphia, PA: Elsevier-Saunders, 2015

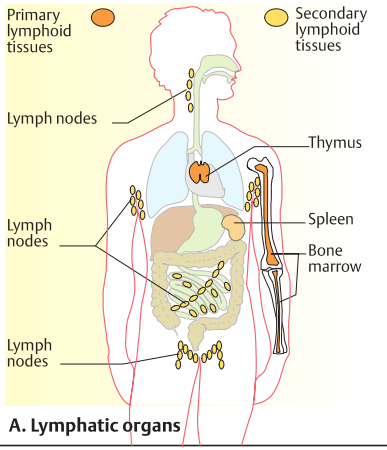
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Haynes BF, et al. Introduction to the immune system. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 2650–2685

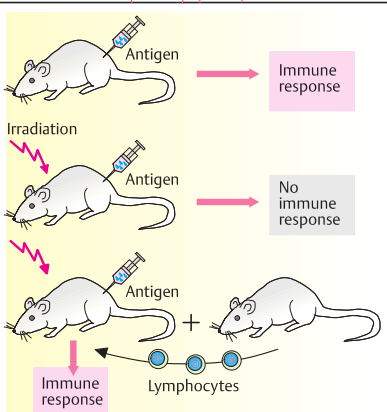
Murphy K, et al. Janeway's Immunobiology. 8th ed. New York, NY: Garland Science, 2011.

Netea MG, et al. Trained immunity: A program of innate immune memory in health and disease. Science 2016; 352(6284): aaf1098

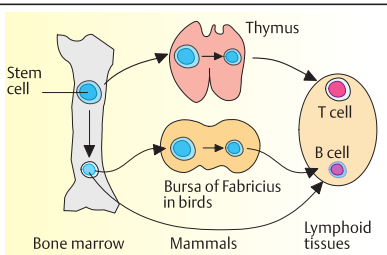
Parham P. The Immune System. 4th ed. New York, NY: Garland Science, 2014



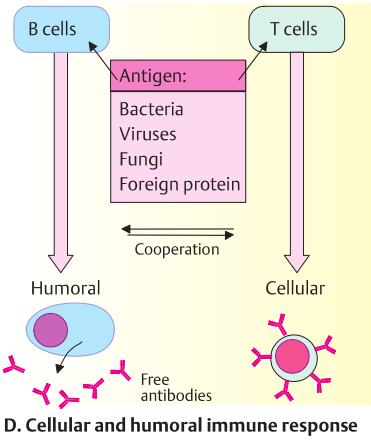
A. Lymphatic organs



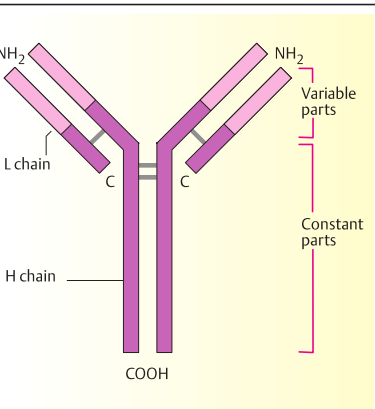
B. Lymphocytes and immune response



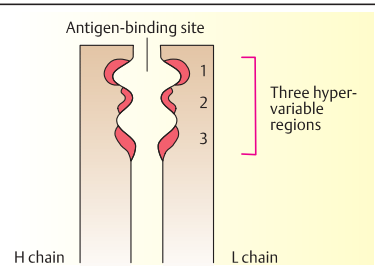
C. T cells and B cells



D. Cellular and humoral immune response



E. Antibody molecule (basic structure)



F. Antigen-antibody binding sites

Immunoglobulin Molecules

Immunoglobulins are the effector molecules of the immune system. They exist in two basic forms: as membrane-bound cell surface receptor molecules or as free antibodies, each in a vast array of variants. Each molecule has two binding sites for a foreign molecule (the antigen). Each binding site contains regions that vary in their amino acid sequence among individual immunoglobulins. This enables each molecule to bind a foreign protein specifically to a particular epitope. A vast number of different effector molecules provide spectacular diversity. Although they differ in details of their structure and function, they share a relatively simple basic pattern, which is derived from a common ancestral molecule during evolution.

A. Immunoglobulin G (IgG)

Immunoglobulin G is the prototype of secreted antibody molecules in humoral immunity. The molecule consists of two H chains and two L chains, held together by disulfide bonds. Each H chain has three constant regions (domains CH1, CH2, and CH3) and one variable region (VH) with a total of 440 amino acids (110 in the V region). Each L chain (214 amino acids) has one constant (C_L) and one variable (VL) domain, also with 110 amino acids. The variable regions of each chain contain the antigen-binding sites with three hypervariable regions called complementarity determining regions (CDRs), where the actual physical contact with foreign epitopes takes place. A region called the hinge joins the constant region 1 (C_H1) and constant region 2 (C_H2) of the heavy chains. This allows considerable flexibility of the molecule. The two H chains, and the H and the L chains, are held together by interchain disulfide bonds (—S—S—). Furthermore, there are intrachain disulfide bonds within each of the polypeptide chains. The L chains are of one of two types: kappa (K) or lambda (λ).

Five classes of immunoglobulin exist: IgA, IgD, IgE, IgG, and IgM. They differ from each other in the constant part of their H chain (C_α, C_δ, C_ε, C_γ). The largest secreted immunoglobulin, IgM, exists as a pentamer of five Ig molecules held together by disulfide bonds. The different types of H chains are referred to as isotypes.

Two proteolytic enzymes cleave Ig molecules into characteristic fragments (**1**). Papain cleaves at the amino terminal disulfide bond (—S—S—) to produce three pieces: two Fab fragments (fragment antigen binding) and one Fc fragment (fragment crystallizable). Pepsin cleaves Ig into one $F(ab')_2$ fragment with an intact disulfide bridge, and one Fc fragment (pFc'), and several small pieces.

B. Structure of an immunoglobulin

Three globular domains of similar size form the **Y**-shaped antibody molecule (immunoglobulin, Ig). The three regions, called domains, are connected by a flexible tether, called a hinge. The two heavy (H) chains are shown in yellow and grey, and the two light (L) chains in brown. The two antigen-binding sites are the two ends of the Y. (Figure adapted from Murphy et al., 2011.)

C. Genes encoding different polypeptides of an immunoglobulin

Each immunoglobulin and receptor molecule is encoded by a different DNA sequence; these sequences belong to a multigene family. The genes for the H chain are located on chromosome 14 at q32 in humans and chromosome 14 in mice. The genes for the **K** light chain are located on chromosome 2 at p12 in humans and chromosome 6 in mice. The genes for the λ light chain are located on chromosome 22 at q11 in humans and on chromosome 16 in mice.

Further Reading

Abbas AK, et al. Cellular and Molecular Immunology. 8th ed. Philadelphia, PA: Elsevier-Saunders, 2015

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

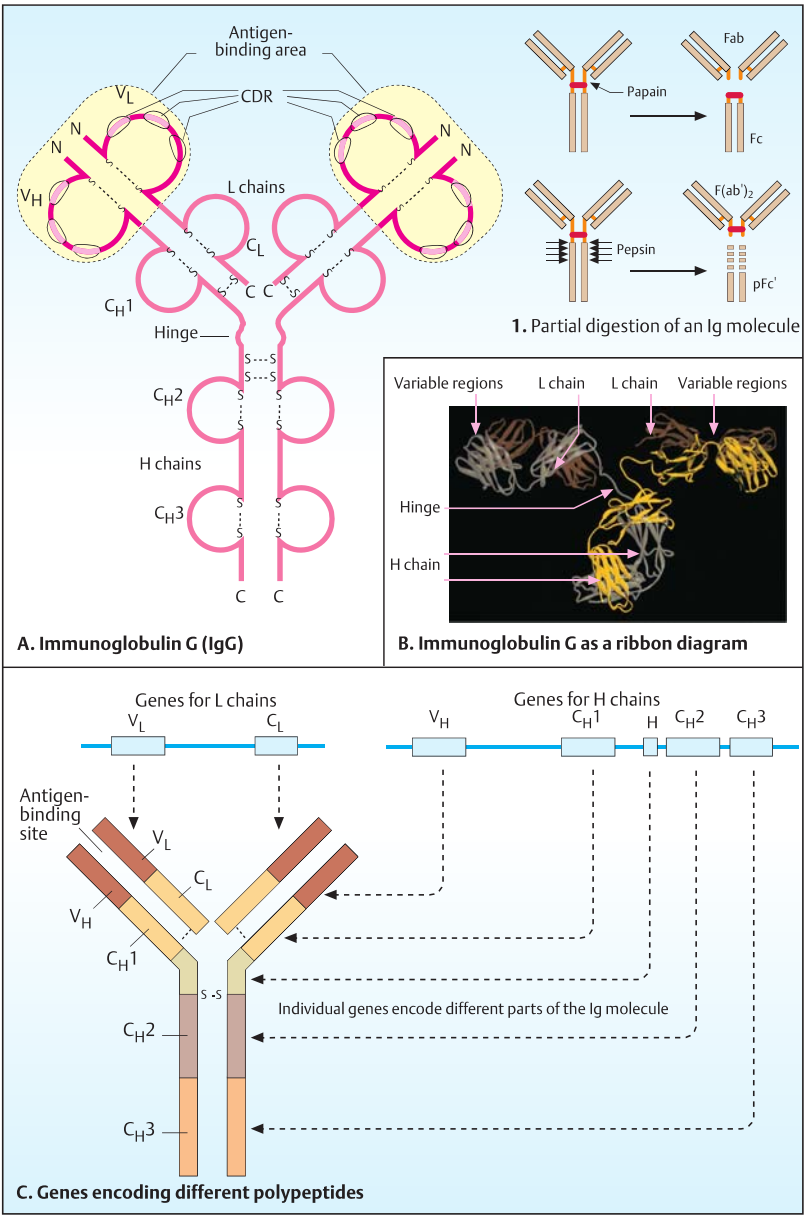
Delves PJ, et al. The immune system. First of two parts. N Engl J Med 2000; 343(1): 37–49, 108–117

Haynes BF, et al. Introduction to the immune system. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 2650–2685

Murphy K, et al. Janeway's Immunobiology. 8th ed. London: Taylor & Francis, 2011

Nossal GJ. The double helix and immunology. Nature 2003; 421(6921): 440–444

Strominger JL. Developmental biology of T cell receptors. Science 1989; 244(4907): 943–950


Antigen-binding area
CDR
V_L
V_H
N
L chains
C_L
C_H1
Hinge
C_H2
H chains
C_H3
C
Papain
Fab
Fc
Pepsin
$F(ab')_2$
pFc'
1. Partial digestion of an Ig molecule
A. Immunoglobulin G (IgG)
Variable regions
L chain
Hinge
H chain
B. Immunoglobulin G as a ribbon diagram
Genes for L chains
Genes for H chains
H
Antigen-binding site
Individual genes encode different parts of the Ig molecule
C. Genes encoding different polypeptides

Generation of Antibody Diversity

Different antibody molecules are generated in response to foreign molecules (antigens) by rearranging DNA segments encoding different parts of each immune molecule. During differentiation of B and T lymphocytes, different segments of genes present in the germline are randomly selected and arranged in a new combination that is specific for each cell and its progeny. This generates a great diversity of cells, each expressing a different antigen-binding molecule (approximately 10^{12} different molecules). Each cell expresses only one particular type of receptor. This is called *allelic exclusion*. The main mechanism is somatic recombination during the differentiation of B cells and T cells. For each domain of an Ig molecule, several selectable sequences are present in genomic DNA. In lymphocyte DNA, these are combined in a variety of combinations for each molecule. Functional genes are produced by gene rearrangements of different elements (V, D, J) from the germline DNA. For each element, several selectable sequences are present. These sequences are combined in a variety of combinations for each gene, encoding one of the polypeptide chains of the T cell or B cell receptor. In addition, somatic mutations occur in the hypervariable regions, leading to further genetic differences.

A. Organization of immunoglobulin loci in the human genome

Both B cells and T cells pass through a series of defined steps of differentiation to reach their final stage. During this process, somatic recombination occurs within the gene loci encoding the two L chains and the H chain. The corresponding loci consist of coding sequences for each of the following segments: V (variable), J (joining), and C (constant). For the H chain, there are in addition 25D (diversity) segments (D_H). The number of functional gene segments is approximately $30V_\lambda$, $40V_\kappa$, and $50V_H$. These gene segments are rearranged during B cell differentiation as shown in **B** for the heavy (H) chain. The C_H genes form a large cluster spanning 200 kb in the 3′ direction from the J segments. Each segment has a leader or signal sequence (L) for guiding the emerging polypeptides into the lumen of the endoplasmic reticulum. (Figure adapted from Murphy et al., 2011.)

B. Somatic recombination during differentiation of lymphocytes

Here, the rearrangements by somatic recombination are shown for the locus of the immunoglobulin H chain (**1**). The first rearrangement joins D and J segments in lymphocyte DNA (D–J joining, **2**). The next rearrangement brings one of the VH genes together with the D–J segments joined previously, resulting in V–D–J joining (**3**). The result is a transcription unit consisting of one V_H, one D_H, one J, and one C gene arranged in this order in the 5′ to 3′ direction (**4**). Each C segment consists of different exons, corresponding to the domains of the complete C region and different isotypes (C_λ, C_δ), etc. The primary transcript is spliced to yield an mRNA consisting of one V, one D, one J, and one C segment (**5**). This is translated into a polypeptide corresponding to the complete H chain (**6**). Following posttranslational modifications, such as glycosylation, the final H chain is produced (**7**).

The result is an array of cells, each with a unique combination of molecules in the antigen-binding site. This provides each molecule with an antigen-binding specificity that differs from that of all other cells. The L chains and the genes encoding the T-cell receptor (see p. 304) are formed in a similar manner. In contrast to the H chains, the L chains have no diversity (D) genes, so that a J gene and a V gene are directly joined by somatic recombination during DNA rearrangement in the lymphocytes. (Figure adapted from Abbas et al., 2015.)

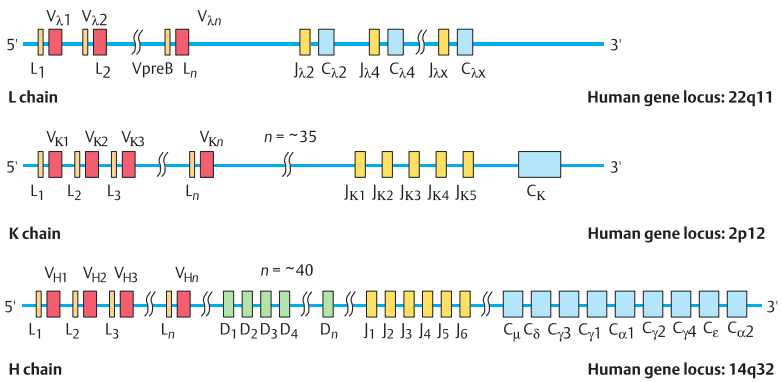
Further Reading

Abbas AK, et al. Cellular and Molecular Immunology. 8th ed. Philadelphia, PA: Elsevier-Saunders, 2015

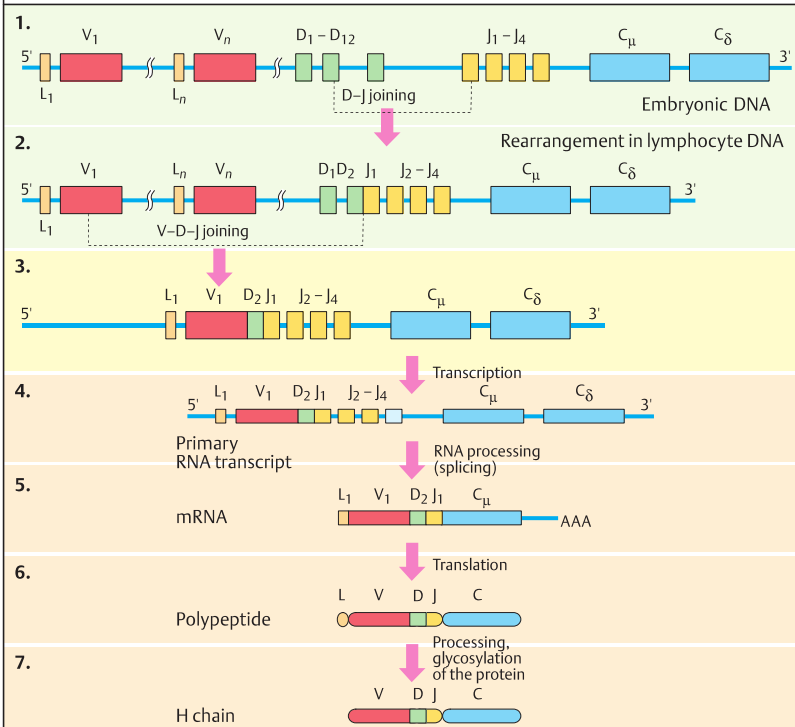
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Murphy K, et al. Janeway's Immunobiology. 8th ed. London: Taylor & Francis, 2011

Schwartz RS. Shattuck lecture: diversity of the immune repertoire and immunoregulation. N Engl J Med 2003; 348(11): 1017–1026



A. Organization of immunoglobulin loci in the human genome



B. Somatic recombination during synthesis of immunoglobulin molecules

Immunoglobulin Gene Rearrangement

The lymphocyte-specific rearrangement of coding DNA segments for immunoglobulins and T-cell receptors follows strict rules under the control of recombination-activating genes (*RAG*1, OMIM 179615; *RAG*2, OMIM 179616). It must be assured that a V gene segment joins a D or J segment, and not another V segment. Conserved noncoding DNA sequences located adjacent to the point of recombination guide this process. The recombinations of V, D, and J segments are performed by lymphocyte-specific DNA-modifying enzymes called V(D)J recombinase. The lymphoid-specific genes *RAG*1 and *RAG*2 are expressed in pre-B cells and immature T cells. The other enzymes involved are DNA-modifying proteins required for double-stranded DNA repair, DNA bending, and ligation of the ends of broken DNA.

A. DNA recognition sequences

The recognition sequences for the *RAG*1 and *RAG*2 genes are noncoding, but highly conserved DNA segments of 7 bp (a CACAGTG heptamer) and a 9-bp nonamer (ACAAAAACC). Recombination takes place between segments located on the same chromosome. A gene segment flanked by a recombination signal sequence with a 12-bp spacer can only be joined to one flanked by a 23-bp spacer. This is referred to as the 12/23 rule. As a consequence, a D segment with a 12-bp spacer on both sides must be joined to a heavy chain J segment. Likewise, a heavy chain V segment can be joined only to a D segment, but not to a J segment, because V and J segments are both flanked by 23-bp spacers. The RAG proteins induce DNA double breaks at the flanking sequences. Rejoining is mediated by DNA repair enzymes (see p. 80).

During formation of an H chain, nonhomologous pairing of the heptamer of a D segment and the heptamer of a J segment occurs. Lymphocyte-specific recombinase enzymes, RAG-1 and RAG-2, recognize DNA sequences located at the 3′ end of each V exon and the 5′ end of each J exon. The RAG-1 and RAG-2 enzymes cleave both strands of DNA at the recognition sites. The RAG proteins align the recognition site sequences and the endonuclease of the RAG complex cuts both strands of the DNA at the 5′ end. This creates a DNA hairpin in the gene segment-coding region. The D and J segments are then joined (D–J joining) by means of recombination: the spacer of 12 or 23 bp and all the intervening DNA form a loop. This loop is excised, and the D and J segments are joined by the nonhomologous end-joining machinery. By pairing and recombination of the recognition sequences at the 5′ end of a D–J segment and the recognition sequence at the 3′ end of a V gene, a V segment is joined to the D–J segment. Diversity of T-cell receptors (see p. 304) is generated in the same manner. (Figure adapted from Abbas et al., 2015.)

B. Antigen genetic diversity

The total diversity, of approximately 10^{18} possible combinations for all types of immunoglobulins and T-cell receptor genes, is the result of different mechanisms. To begin with, different numbers of variable DNA segments are available for different chains (250–1,000 for the H chain, 250 for the L chains, 75 for the α chain of the T-cell receptor TCRα, etc.). The different D and J segments also multiply the number of possible combinations. Finally, DNA sequence changes (somatic mutations) occur regularly in the hypervariable regions, further increasing the total number of possible combinations.

Medical relevance

Mutations in the RAG1 and RAG2 genes (OMIM 179615/16) cause severe combined immune deficiency (SCID; OMIM 600802. see p. 310) as well as Omenn syndrome, which is caused by defective V(D)J recombination.

Further Reading

Abbas AK, et al. Cellular and Molecular Immunology. 8th ed. Philadelphia, PA: Elsevier-Saunders, 2015

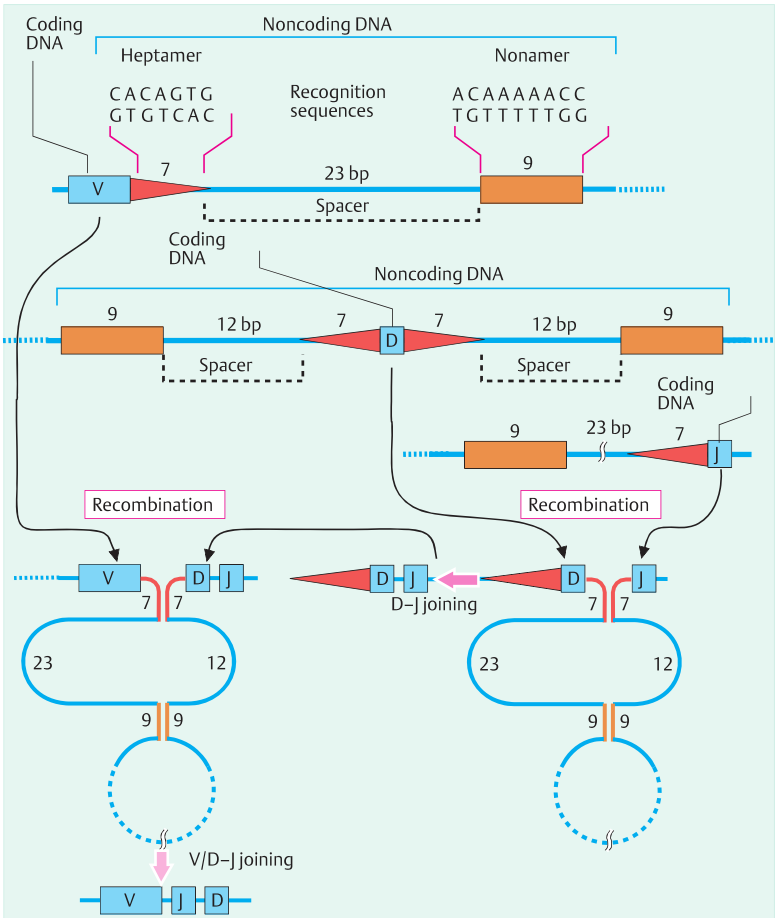
Agrawal A, et al. RAG1 and RAG2 form a stable post-cleavage synaptic complex with DNA containing signal ends in V(D)J recombination. Cell 1997; 89 (1): 43–53

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Matthews AGW, et al. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature 2007; 450(7172): 1106–1110

OMIM (Online Mendelian Inheritance of Man). Available at: www.ncbi.nlm.nih.gov/omim

Schwarz K, et al. RAG mutations in human B cell-negative SCID. Science 1996; 274(5284): 97–99



A. DNA recognition sequences for lymphoid gene rearrangement

Mechanism	Immunoglobulin		TCRαβ		TCRγδ	
	H chain	L chain	α	β	γ	δ
Variable domain	250–1000	250	75	25	7	10
Number of D segments	12	0	0	2	0	2
Number of J segments	4	4	50	12	2	2
Variable segment combinations	65 000–250 000		1825		70	
Total diversity	10^{11}		10^{16}		10^{18}	

B. Genetic diversity in immunoglobulin and T-cell receptor genes

T-cell Receptor

T lymphocytes have a membrane-bound protein, the T cell receptor (TCR). The TCR enables T cells to recognize antigens displayed on the surface of a cell from their own body. The antigens are small peptides derived from viruses or intracellular bacteria. They are presented to the TCR by the MHC class I and class II molecules (see next page). The genes encoding the α/β and δ TCRs are arranged in germline DNA according to the segments they encode, that is, variable segments (V), diversity (D), joining (J), and constant regions (C). Gene segments are rearranged during maturation in the thymus in the same manner as the genes encoding the immunoglobulins.

A. T-cell receptor structure

The T-cell receptor (TCR) resembles the Fab fragment of an immunoglobulin molecule. It is a heterodimer of one α and one β polypeptide chain covalently linked by a disulfide bridge and produced as an integral membrane protein (**1**). The basic structure is similar to that of the cell surface immunoglobulins. A subtype of TCR consists of a γ chain and a δ chain. The three-dimensional structure (**2**) reveals that the hypervariable regions CDR1, CDR2, and CDR3 can be aligned with antigen-binding sites of the antibodies. Genomic DNA contains genes for 50 to 70 V segments, 2 D segments, 12 to 60 J segments, and 2 C segments. In a given T cell, only one of the two parental genes for a given α chain and β chain is expressed (allelic exclusion). The β chain is the slightly larger chain. The V region of each chain consists of 102 to 109 amino acids and contains three hypervariable regions, as do the immunoglobulin molecules. Genes for T-cell receptor γ and δ chains exist in addition to those for α and β chains. (Image in 2 adapted from Murphy et al., 2011.)

B. Interaction of T-cell receptor and MHC

The TCR interacts with an antigen-presenting cell carrying a major histocompatibility complex (MHC) molecule, either a human leukocyte antigen (HLA) class I or a class II molecule (**1**). The three-dimensional structure (**2**) shows the tight connection between the interacting molecules. T cells able to destroy an infected cell (cytolytic T lymphocytes, CTLs, or "killer cells") recognize their antigen on MHC class I molecules, while "helper cell" T cells specifically bind to MHC class II molecules. CD8 cells are restricted to MHC class I molecules, and CD4 cells are restricted to MHC class II molecules. (Image in 2 adapted from Murphy et al., 2011.)

C. Recognition of antigen and T-cell activation

A few of the many molecules involved in T-cell activation are shown. CD4 and CD8 molecules serve as restriction elements. By binding directly to the MHC class II molecule, CD4 stabilizes the TCR interaction with the peptide antigen. CD8 takes this role with cytolytic T lymphocytes (CTLs, "killer cells") binding to MHC class I molecules. Upon antigen recognition by the TCR, the associated CD3 complex is phosphorylated. Complete T-cell activation requires the engagement of costimulatory receptors (CD28, LFA-1) on T cells and their ligands (B-7 and ICAM-1) on antigen-presenting cells. This signal transduction activates the *IL2* (interleukin 2) gene. Interleukin 2 is the main T-cell growth factor of T cells and is responsible for progression in the cell cycle from the G1 to the S phase.

Medical relevance

The gp120 protein of the HIV virus interacts with the second domain of CD4 cells.

Further Reading

Abbas AK, et al. Cellular and Molecular Immunology. 8th ed. Philadelphia, PA: Elsevier-Saunders, 2015

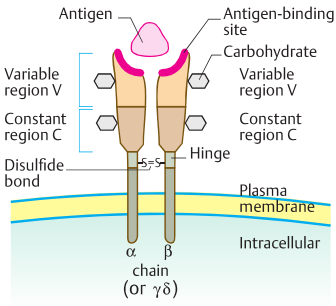
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Amadou C, et al. Localization of new genes and markers to the distal part of the human major histocompatibility complex (MHC) region and comparison with the mouse: new insights into the evolution of mammalian genomes. Genomics 1995; 26(1): 9–20

Fugger L, et al. The role of human major histocompatibility complex (HLA) genes in disease. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001

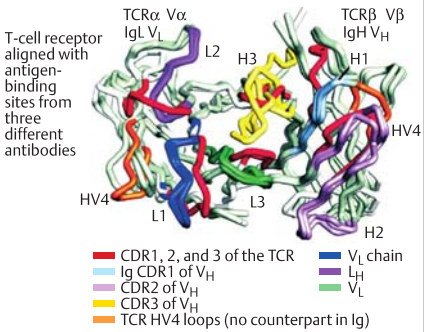
Jiang H, et al. Regulation of immune responses by T cells. N Engl J Med 2006; 354(11): 1166–1176

Murphy K, et al. Janeway's Immunobiology. 8th ed. London: Taylor & Francis, 2011

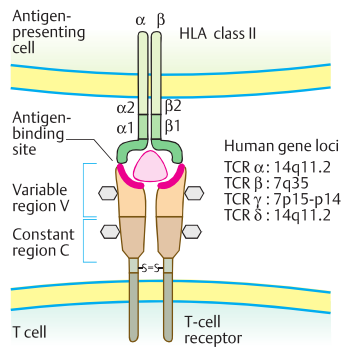


1. Schematic structure

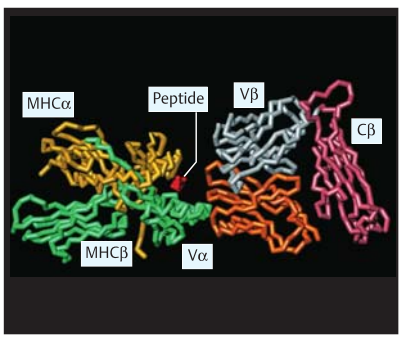
A. T-cell receptor



2. Crystal structure at 2.5 Å

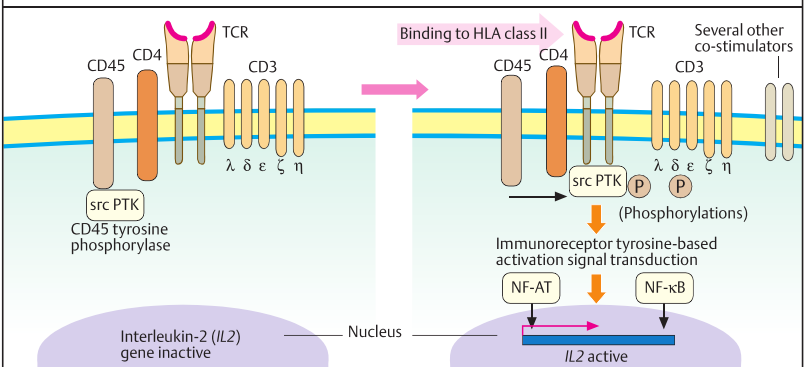


1. Schematic structure



2. Three-dimensional structure

B. Interaction of T-cell receptor and MHC class I or class II molecules



C. Recognition of antigen and T-cell activation

The MHC Region

The major histocompatibility complex (MHC, OMIM 613503), usually referred to as the human leukocyte antigen (HLA) system (human leukocyte antigen), is a large chromosomal region containing more than 400 genes along nearly 4,000 kb (4 megabases, Mb) of DNA with the highest density of genes in the human genome (6 per 100 kb, average 1). It is the most polymorphic region with up to 30 to 60 alleles at many loci, with more than 400 alleles at HLA-A, more than 700 at HLA-B, and more than 500 at HLA-DRB1. Each MHC molecule is composed of two polypeptide chains and is expressed on the surface of various cells. This can be demonstrated serologically by a binding assay or by a cytotoxic cellular reaction in a mixed lymphocyte test.

A. Genomic organization of the MHC complex

In human, the MHC region is located on the short arm of human chromosome 6 (6p21.32); in the mouse, it is located on chromosome 17. The genes are grouped into three classes: I to III. Class I and class II belong to the HLA system; in mice it is called the H2 system. Class I in humans consists of HLA-A, HLA-B, HLA-C, and several other loci (HLA-E to HLA-J). Mouse class I genes form two groups: D and L at the 3′ end, and K at the 5′ end. Human class II MHC molecules are grouped into DP, DQ, and DR (in mice, they are I-A and I-E; letter I, not the roman numeral). Class III includes genes that are not directly involved in the immune system as well as genes involved in the immune response, such as tumor necrosis factor (TFN), and lymphotoxin.

B. Gene loci in the MHC

The subgroups DP, DQ, and DR of the human class II HLA loci, oriented toward the centromere, are subdivided according to their composition of α chains and β chains (some genes between DM and DO are not shown). The DR locus contains three β chains, each of which can pair with the single α chain. Thus, three sets of genes can produce four types of DR molecules. The class III loci, located between class I and class II, contain genes encoding complement factors C2 and C4 (*C4A*, *C4B*), steroid 21-hydroxylase (*CYP21B*), and cytokines (tumor necrosis factor, *TNFA*; lymphotoxin, *LTA*, *LTB*). Several other class III genes are located in the class III region (not shown). The three-dimensional structure shows the tight arrangement of polypeptide chains of the MHC class II protein. (Image adapted from Murphy et al., 2011.)

C. Structure of MHC molecules

The MHC class I and class II molecules have distinct structures. Class I molecules, HLA-A, HLA-B, and HLA-C, consist of single membrane-bound polypeptides, an α chain with three domains: α1 at the extracellular N-terminal, α2, and α3 (**1**). The α chain is noncovalently bound to nonpolymorphic β_2-microglobulin, which is encoded by a gene on human chromosome 15q13-q15 (OMIM 109710). The three extracellular domains of the α chain have approximately 90 amino acids each. The α1 and α2 regions form the highly polymorphic peptide-binding region. The α3 domain and β_2-microglobulin structurally correspond to an immunoglobulinlike region. Class II MHC molecules (**2**) have two polypeptide chains, α and β, each with two domains, α1 and α2, and β1 and β2; each has 90 amino acids and a transmembrane region of 25 amino acids. The peptide-binding regions α1 and β1 are highly polymorphic. The crystalline structures of the MHC molecules have revealed details of the binding mechanisms (**3**). (Figure adapted from Bjorkman et al., 1987.)

Medical relevance

Certain allelic variants predispose to disorders associated with an altered immune response or immune deficiency.

Further Reading

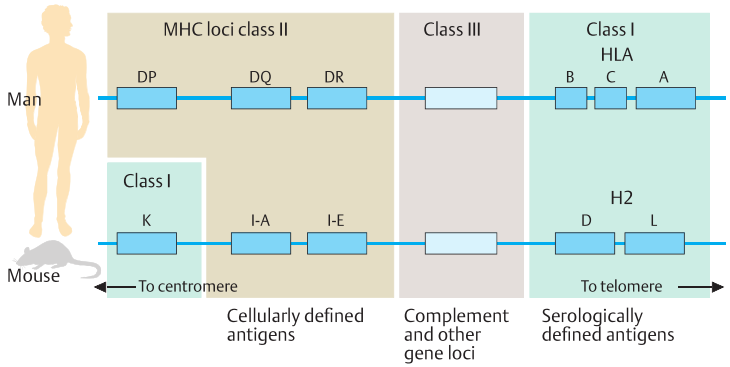
Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Bjorkman PJ, et al. Structure of the human class I histocompatibility antigen, HLA-A2. Nature 1987; 329(6139): 506–512

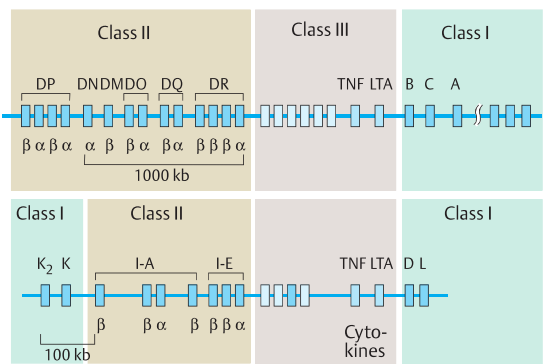
Murphy K, Travers P, Walport M. Janeway's Immunobiology. 8th ed. New York: Taylor & Francis, 2011

Nepom GT. The major histocompatibility complex. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012

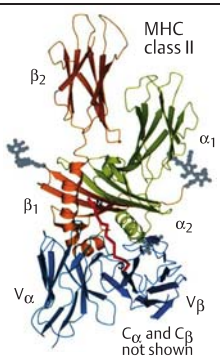
OMIM www.ncbi.nlm.nih.gov/omim



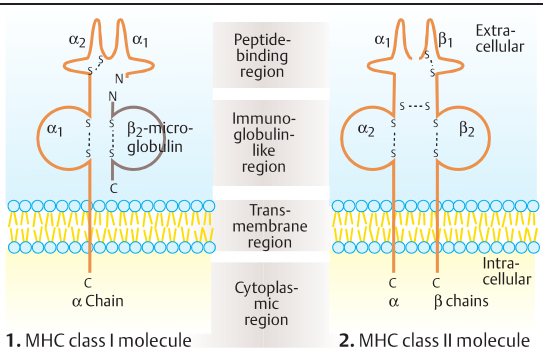
A. Basic structure of the MHC gene complex in man and mouse



B. Genomic organization of the MHC loci

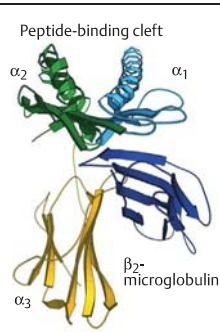


Three-dimensional structure



1. MHC class I molecule

2. MHC class II molecule



3. Ribbon diagram of an MHC class I molecule

C. MHC molecules of class I and class II

Evolution of the Immunoglobulin Superfamily

The structural similarity of most cell–cell or cell–antigen recognition molecules of the immune system suggests that they share a common evolutionary history, forming the immunoglobulin (Ig) superfamily. A prominent functional feature is their ability to bind other molecules with high specificity. The common structural leitmotif is the **Y**-shaped Ig structure. A prominent attribute of the Ig proteins is their ability to adhere directly to another cell or a foreign protein through specialized cell membrane proteins known as the cell adhesion molecules (CAMs). This large family of molecules is classified into four major families: the Ig superfamily, cadherins, integrins, and selectins. CAMs are usually built up in repeats of domains with different properties. The Ig superfamily comprises a large group of molecules with globular domains. A gene superfamily is a collection of genes of common evolutionary origin that arose by gene duplications and subsequently diverged into genes with new, different functions.

A. Basic structure of proteins of the immunoglobulin supergene family

Characteristic structures shared by Igs are repeated domains, usually of approximately 70 to 110 amino acids with variable (V) and constant (C) domains (**1**). Each Ig domain is derived from conserved DNA sequences. The prototype, IgG, has three C domains in the heavy chain, one C domain in each of the two light chains, and a variable domain in each chain. The Ig molecules of the T-cell receptors (TCRs) and the class I and class II MHC molecules are basically similar. Although their genes are located on different chromosomes, the gene products form functional complexes with each other. Others, such as the V, D, and J gene segments of all antigen receptors and their genes for the C domain, form gene clusters. Genes of the MHC loci and for the two CD8 chains lie together.

Accessory molecules such as CD2, CD3, CD4, CD8, and thymosin 1 (Thy-1) are members of this family with a relatively simple but similar structure (**2**). Other members of the Ig superfamily are cell adhesion molecules, such as the Fc receptor II (FcRII); the polyimmunoglobulin receptor (pIgR), which transports antibodies through the membranes of epithelial cells; NCAM (neural cell adhesion molecules); and PDGFR (platelet-derived growth factor receptor, 3). (Figure adapted from Hunkapiller and Hood, 1989, and Alberts et al., 2008.)

B. Evolution of genes of the immunoglobulin supergene family

Distinct evolutionary relationships can be recognized by the homology of genes for Ig-like molecules and their gene products. A precursor gene for a variant (V) region and a constant (C) region must have arisen from a primordial cell by duplication, and then subsequently diverged into different cell surface receptor genes. Further duplication events have resulted in structurally related gene segments encoding proteins with repeated domain structures. At an early stage, rearrangements between different gene segments occurred and became the standard for Igs, T-cell receptors, and CD8. Other members of this superfamily evolved into cell adhesion molecules without somatic recombination, such as the thymosin (Thy-1) receptor or the poly-Ig receptor. It is clear that somatic recombination of the genes for antigen-binding molecules has enormous evolutionary advantages. (Figure adapted from Hood et al., 1985.)

Further Reading

Abbas AK, et al. Cellular and Molecular Immunology. 8th ed. Philadelphia, PA: Elsevier-Saunders, 2015

Alberts B, et al. Molecular Biology of the Cell. 5th ed. New York, NY: Garland Science, 2008

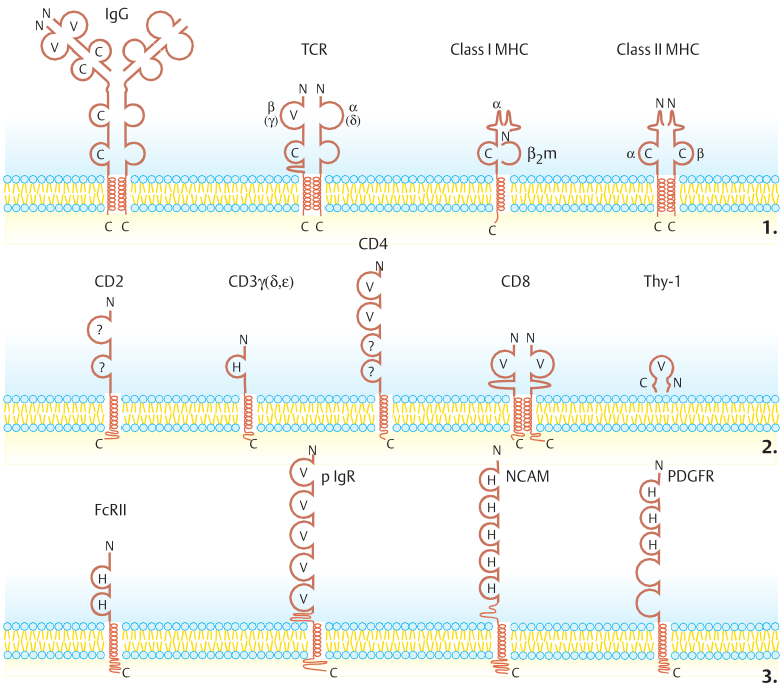
Hood L, et al. T cell antigen receptors and the immunoglobulin supergene family. Cell 1985; 40(2): 225–229

Hunkapiller T, Hood L. Diversity of the immunoglobulin gene superfamily. Adv Immunol 1989; 44: 1–63

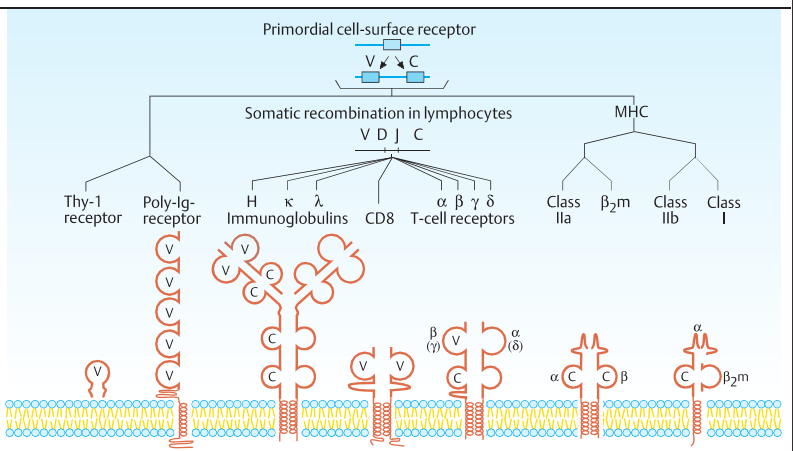
Klein J, Sato A. The HLA system. First of two parts. N Engl J Med 2000; 343: 702–709 (part I), 782–786

Klein J, Takahata N. Where Do We Come from? The Molecular Evidence for Human Descent. Heidelberg: Springer-Verlag, 2002

Shiina T, et al. Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region. Proc Natl Acad Sci U S A 1999; 96(23): 13282–13287



A. Basic structure of proteins of the immunoglobulin supergene family



B. Evolution of genes of the immunoglobulin supergene family

Primary Immunodeficiency Diseases

Primary immunodeficiency diseases comprise a group of approximately 200 genetic disorders resulting from mutations in approximately 150 different genes with a prevalence of approximately 1 in 20,000 individuals. They often result in severe, life-threatening diseases. Genetic immunodeficiency occurs either isolated or as manifestations of a multisystem disease. Primary immunodeficiency disorders involve adaptive and innate immunity.

A. Hereditary immunodeficiency diseases: overview

This group of diseases can be classified according to which branch of the immune system is primarily or solely involved. One group, severe combined immune deficiency (SCID, OMIM 601457), results from a genetic block of differentiation of precursor cells before they differentiate into B cells or T cells. Thus, both cell types are involved. It is a heterogeneous group of genetic disorders resulting from various defects in both B-cell and T-cell differentiation.

X-linked agammaglobulinemia type Bruton (300300) was the first hereditary immune deficiency described in 1952 by Ogden Bruton (see **B**). Here, the first developmental step of B-cell differentiation from pre-B to mature B cell is blocked by deficiency of Bruton tyrosine kinase (Btk) as a result of mutations in the *BTK* gene on the X chromosome (Xq22.1). This gene is the key regulator of B-cell development. The immunodeficiency results in failure of B lymphocytes to mature and failure of Ig heavy chain rearrangements. OMIM 601495 lists other forms of agammaglobulinemia due to mutations in 10 genes with a known molecular basis.

Dysregulated later steps of differentiation lead to variable immune deficiency or isolated Ig isotype (subclass) deficiencies. Several immune deficiency diseases involve T cells, such as T-cell receptor (TCR) signal transduction, V(D)J recombination (mutations in *RAG*1 and *RAG*2 genes; OMIM 179615/16), cytokine signal transduction, and apoptosis regulation. Several disorders involve T-cell activation and the function of one or both major subsets of T cells: CD4 or CD8. Predisposition to tumors of the lymphoid system and autoimmune dysfunction is relatively frequent in immunodeficiency diseases. Effective therapy by bone marrow transplantation is possible in some of the diseases.

B. Severe combined immune deficiency

SCID is a heterogeneous group of disorders. The most common is X-linked SCID (OMIM 308380, 300400). SCIDX1 (300400) results from mutations in the X-chromosomal gene IL2RG on Xq13.1, encoding the γ subunit of the interleukin 2 receptor, gamma-C (IL2Rγ; 308380). This subunit is shared with other interleukin receptors. IL2RG is an atypical member of the cytokine receptor family. More than 200 different mutations have been recorded. (Figure adapted from Burmester and Pezzuto, 2003.)

C. DiGeorge syndrome and deletion 22q11

DiGeorge syndrome (188400) is a combination of T-cell defects associated with a highly variable spectrum of congenital malformations involving absence of the thymus and other derivatives of the embryonic third and fourth branchial arches. If the parathyroid glands are absent, neonatal hypocalcemia is immediately life threatening. In addition, abnormalities of the face may be present. DiGeorge syndrome is now considered part of an overlapping spectrum of disorders involving various-sized deletions of 22q11 (microdeletion syndromes). This includes a previously recognized distinct disorder, the velocardiofacial syndrome of Shprintzen (192430). Other examples of immune deficiency diseases are listed in the Appendix, Table 12 p. 412. (Figure adapted from Burmester and Pezzuto, 2003.)

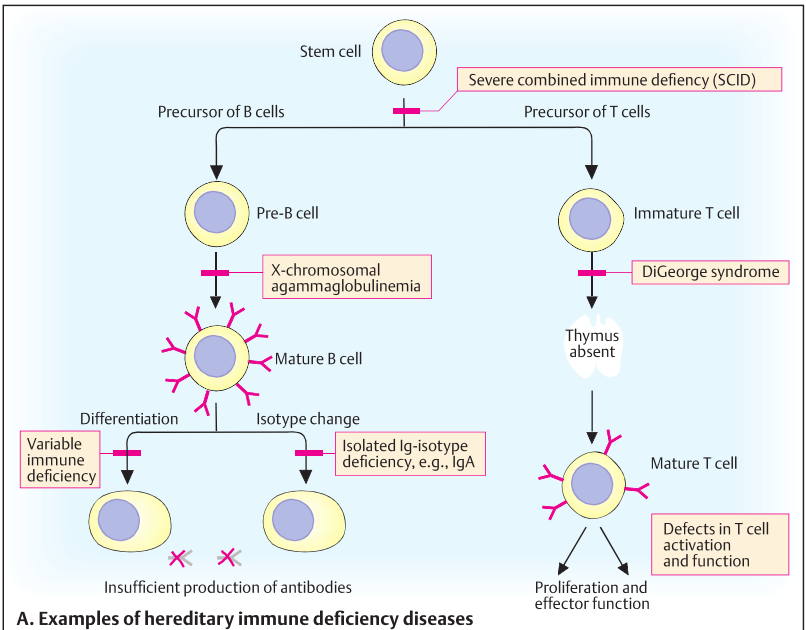
Further Reading

Belmont JW, Puck JM. T cell and combined immunodeficiency disorders. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 4751–4783

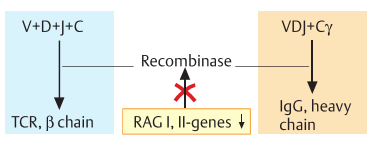
Burmester GR. Pezzuto A. Color Atlas of Immunology. Stuttgart: Thieme, 2003

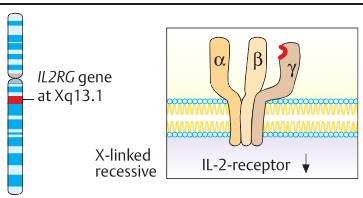
Fischer A. Primary immune deficiency diseases. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012

Notarangelo LD, et al. International Union of Immunological Societies Expert Committee on Primary Immunodeficiencies. Primary immunodeficiencies: 2009 update. J Allergy Clin Immunol 2009; 124(6): 1161–1178

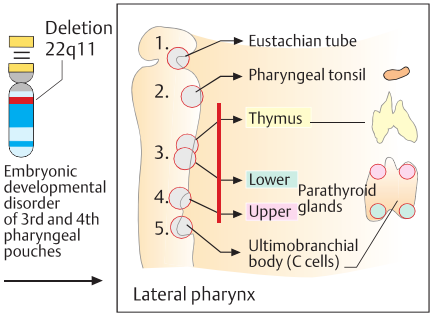


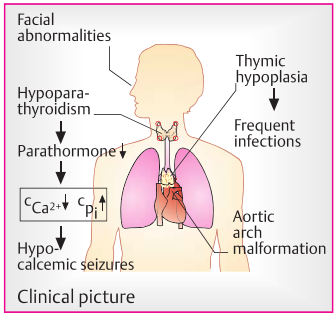
A. Examples of hereditary immune deficiency diseases





B. Severe combined immunodeficiency





C. Di George syndrome (deletion 22q11)

Genetic Causes of Cancer

Cancer is a heterogeneous group of approximately 200 different types of tumors. Each is characterized at the cellular level by unrestrained proliferation of cells transformed to a malignant state. Multiple sequential mutations and genomic rearrangements in specific genes (cancer genes) follow an initial event, as first suspected by Boveri in 1914. Cancer affects one in every five individuals over lifetime. A cancer cell breaks two rules imposed on all cells in a multicellular organism: this cell and its daughter cells do not adhere to restrained cell division and they eventually invade and colonize other tissues. Cancer is classified according to the tissue of origin: *carcinoma* from epithelial cells, *sarcoma* from connective tissue and muscle cells, *leukemia* from hematopoietic cells, and *lymphoma* from lymphoid cells. In 85%, cancer occurs in single individuals due to somatic mutations. In up to 15%, depending on the type of cells affected, a predisposing mutation present in the germline represents a hereditary component and may result in familial clustering.

A. Multistep clonal expansion of malignancy

Starting usually from a single cell, a population of genetically altered cells acquires properties that confer a selective growth advantage over their normal neighbors. Such a cell (**1**) may divide after passing an inhibitory barrier (e.g., cell cycle control). A second change may follow (**2**) and another selective barrier passed (**3**). Eventually, tumor progression ensues (**4**). Dividing cells are prone to errors in replication. If this is not recognized and eliminated, it will be transmitted to daughter cells. Approximately 10^{16} cell divisions occur over an average human life time with an estimated rate of spontaneous mutations of 10^{-6} per gene per cell division. Thus, age is an important contributing factor to the causes of cancer. (Figure based on Nowell, 1976.)

B. Four basic types of genetic alteration in tumor cells

The genetic alterations affecting growth-controlling genes can be classified into four major categories. (**1**) *Somatic mutation*: in the example shown, a deletion of two adenines (A) in position 128 is present in the DNA sequence of a growth-controlling gene, *TGFBR2* (receptor type 2 of the transforming growth factor beta). In a colorectal cancer (CRC) cell line, this changes the codon AAG (lysine) to GCC (alanine), followed by TGG (tryptophan) and a stop codon (TGA), resulting in a truncated protein. (**2**) *Reciprocal chromosome translocation*: the breakpoint of a translocation between a chromosome 1 and a chromosome 17 disrupts one or two genes as a cause of neuroblastoma (OMIM 256700) cell line. (**3**) *Loss or gain of a chromosome*: a chromosome 3 and a chromosome 12 (yellow arrows) are lost in a clone of a colorectal cell line (SW837). (**4**) *Amplification* of chromosomal regions and the genes they carry: here the chromosomal region contains the N-*myc* gene (OMIM 164840) shown in yellow. It occurs in 40% of childhood neuroblastoma (OMIM 256700) and indicates a poor prognosis. Amplifications were first visualized as homogeneously stained regions (HSRs) in 1976. In addition, epigenetic changes or catastrophic chromosome breaking events (chromothripsis) have been observed (Stephens et al., 2011, not shown). (Data and figures adapted from Lengauer et al., 1998.)

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. N Engl J Med 2013; 368(9): 842–851

Boveri T. Zur Frage der Entstehung maligner Tumoren. Jena: G. Fischer, 1914

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144(5): 646–674

Lengauer C, et al. Genetic instabilities in human cancers. Nature 1998; 396(6712): 643–649

Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: Freeman, 2016

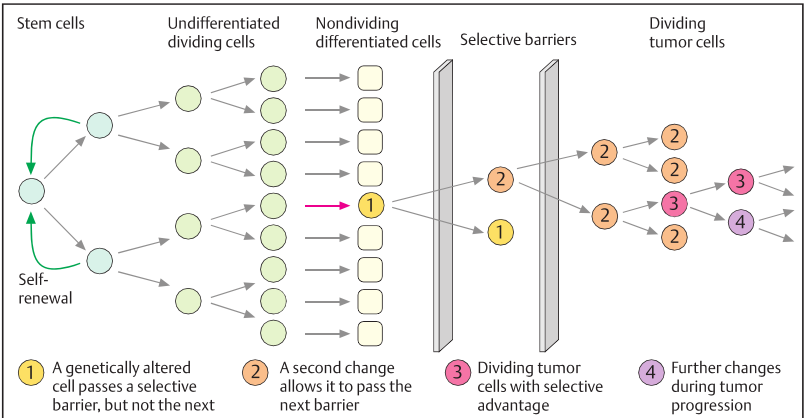
Morin PJ, et al. Cancer genetics. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 663–672

Nowell PC. The clonal evolution of tumor cell populations. Science 1976; 194(4260): 23–28

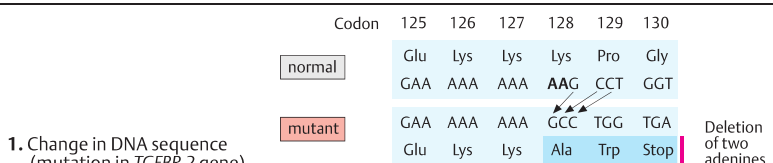
Stephens PJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 2011; 144(1): 27–40

Vogelstein B, et al. Cancer genome landscapes. Science 2013; 339(6127): 1546–1558

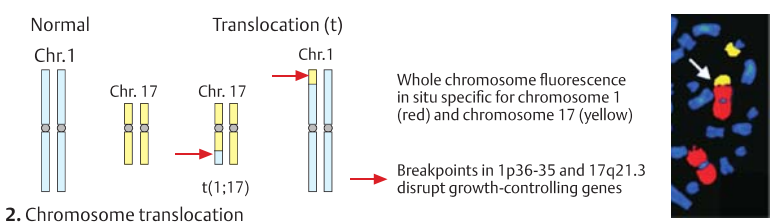
Weinberg RA. The Biology of Cancer. 2nd ed. New York, NY: Garland Science, 2013



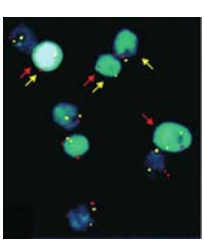
A. Multistep clonal development of malignancy



1. Change in DNA sequence (mutation in *TGFBR 2* gene)

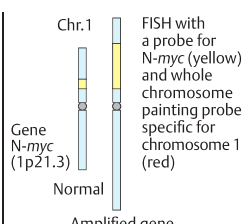


2. Chromosome translocation



Loss of chromosome 3 (red arrows) and chromosome 12 (yellow arrows) in colorectal cancer cells

3. Gross chromosomal change



FISH with a probe for N-*myc* (yellow) and whole chromosome painting probe specific for chromosome 1 (red)

4. Gene amplification

B. Four basic types of genetic alteration in tumor cells

Categories of Cancer Genes

Mutations in approximately 140 genes actively promote tumor development ("driver" mutations). They are classified into three basic categories according to the effects of their mutations: *oncogenes*, *tumor suppressor* genes, and *genome stability* genes.

A. Three categories of cancer genes

The first class is *proto-oncogenes* (**1**). Their mutant forms, called oncogenes, drive a cell to divide when it normally should not (gain-of-function mutation). A single activating mutation is the first step toward cancer (comparable to a stuck accelerator in a car). The second class is *tumor suppressor genes* (**2**), which require two mutational events to induce tumor development (comparable to a defective brake). The initial mutation predisposes the cell to become a cancer cell. The second mutation then inactivates the other allele (loss-of-function mutation) and results in loss of cell division control. Mutations in the third class of cancer genes (**3**), *stability genes*, affect the stability of the genome by disrupting one of the various repair processes.

B. Oncogene activation

An oncogene positively influences tumor formation following a mutation in one allele. Oncogenes serve in many signal pathways controlling cell division. For example, the Ras genes *HRAS*, *KRAS*, and *NRAS* (OMIM 190020) encode a family of related cell-growth-controlling proteins. Ras proteins are GTPase-binding proteins (GAP) functioning as switches, inactive when bound to GDP (guanosyldiphosphate) and active when bound to GTP (guanosyltriphosphate). Ras is activated by a receptor tyrosine kinase, which activates a guanine nucleotide exchange factor (GEF). Mutant forms of Ras are hyperactive and do not respond to GAPs.

C. Tumor suppressor genes

A mutation in a tumor suppressor gene will only initiate tumor formation if both alleles have lost their normal function. Thus, two successive mutational events are required within the same cell (**1**). The first event inactivates one allele, but does not result in a cellular tumor phenotype. However, it predisposes the cell to malignant transformation. This occurs when the other allele is also inactivated by a mutation. Tumor suppressor genes can be assigned to two groups: *gatekeepers* and *caretakers*. Gatekeeper genes directly inhibit tumor growth. Inactivation of caretaker genes leads to genetic instability, indirectly promoting tumor growth.

Loss of one allele in a somatic cell carrying a mutation in the other allele can be visualized by Southern blot analysis (**2**). Whereas somatic cells heterozygous at a marker locus will give two signals, tumor cells which have one allele of the gene only will show one signal (loss of heterozygosity, LOH). LOH is a hallmark of a tumor suppressor gene.

Mutations in tumor suppressor genes may be present in the zygote (by transmission or by new mutation) or occur in a somatic cell (**3**). A germline mutation predisposes all cells to develop into tumor cells. A somatic mutation predisposes a single cell. Germline mutations are the basis for hereditary forms of cancer; somatic mutations are the basis for the nonhereditary forms. A germline mutation occurring after the initial division of the fertilized egg may result in a mosaic of mutated and normal cells.

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

Bozic I, et al. Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci U S A 2010; 107(43): 18545–18550

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144(5): 646–674

Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature 1997; 386(6627): 761–763, 763

Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013; 499(7457): 214–218

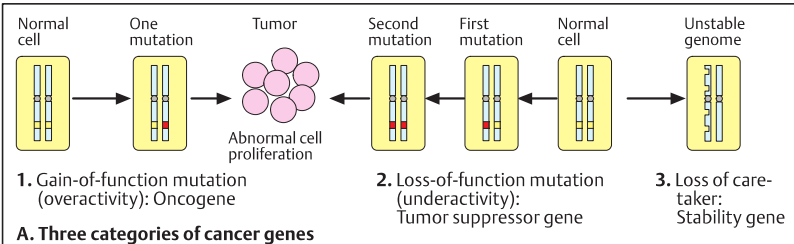
Morin PJ, et al. Cancer genetics. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 663–672

Tokheim CJ, et al. Evaluating the evaluation of cancer driver genes. Proc Natl Acad Sci U S A 2016; 113 (50): 14330–14335

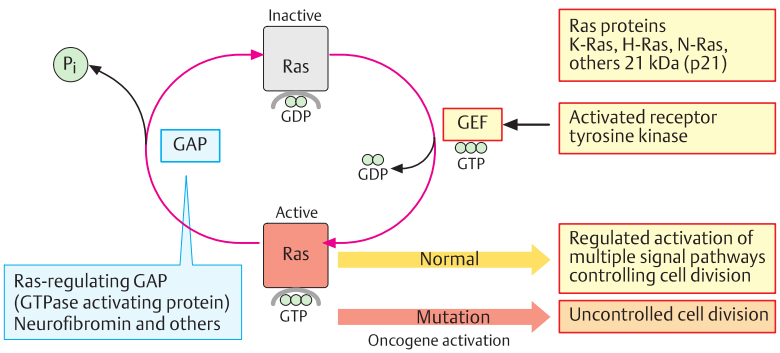
Vogelstein B, et al. The Genetic Basis of Human Cancer. 2nd ed. New York, NY: McGraw-Hill, 2002

Vogelstein B, Kinzler KW. Cancer genome landscapes. Science 2013; 339(6127): 1546–1558

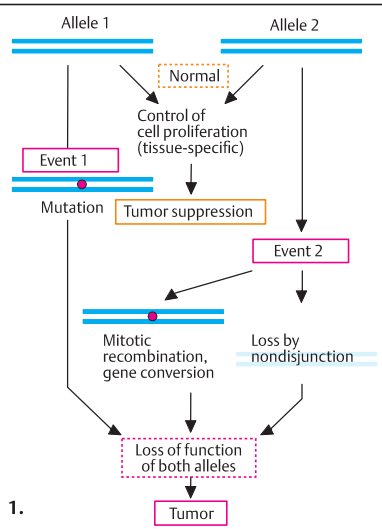
Watson IR, et al. Emerging patterns of somatic mutations in cancer. Nat Rev Genet 2013; 14(10): 703–718



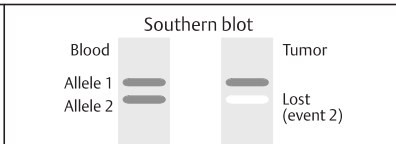
A. Three categories of cancer genes

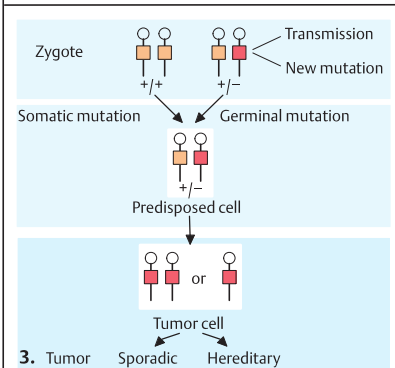


B. Oncogene activation



C. Tumor suppressor gene





Cancer Genomes

More than 30 different tumor types from 11,000 patients have been investigated by different methods including complete sequencing as recorded in The Cancer Genome Atlas. In particular, genomic analysis and massively parallel sequencing (p. 64) has provided insight into the genome of cancer cells. More than 300 genes (cancer genes, 1.8% of the 21,000 protein-coding human genes) have been identified to date by the Cancer Genome Project (Cancer Gene Census). A ring-shaped diagram, called circos plot, is used to display the genome of a cancer cell.

A. Sequence of events from a normal cell to a cancer cell

Age greatly influences the number of mutations accumulating during life time. A normal cell accumulates mutations by an intrinsic process, influenced by lifestyle and environment. Early clonal expansion of cells accumulating mutations precedes the early stages of tumor development (mutator phenotype). Approximately 140 identified driver mutations as part of approximately 12 signaling pathways confer a growth advantage. Cell fate, cell survival, and genome maintenance are regulated in the interacting pathways (see Vogelstein et al., 2013). On average, a tumor growth is initiated by two to eight driver mutations. Additional mutations present do not confer growth advantage ("passenger" mutations). Subsequent stages involve the invasion of surrounding tissues and metastasis. Resistance to chemotherapy marks the final stages of a tumor. (Figure adapted from Stratton et al., 2009.)

B. Circos plot

A circos plot depicts the location of all mutations and other changes in a cancer genome in a circular arrangement. An outer circle shows the chromosomes schematically. In a second circle on the inside, each mutation is indicated at the site of its location with respect to each chromosome. A third circle, located further inside, indicates changes in DNA copy number, either an increase or a decrease of the normal amount of DNA at each site. In the most inner field, chromosomal rearrangements are shown: intrachromosomal rearrangements by a vertical line and interchromosomal rearrangements by a line connecting two chromosomes. Such a cancer genome is compared with the normal genome of an unaffected cell from the same individual. Genome sequences from more than 25,000 have been determined. (Figure adapted from Ledford, 2010, and Stratton et al., 2009.)

C. Genomic profile of an adenocarcinoma

This plot illustrates the key features of the genome of an adenocarcinoma. Red lines indicate interchromosomal structural variations; blue lines represent intrachromosomal structural variations. This figure was created by using the circos program described by Krzywinski et al. in 2009. (Image adapted from Lee et al., 2010.)

Further Reading

Beroukhim R, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010; 463(7283): 899–905

COSMIC (catalogue of somatic mutations in cancer). Available at: www.sanger.ac.uk/genetics/CGP/Census

NIH Cancer Genome Atlas. Available at: https://cancergenome.nih.gov/abouttcga/overview

Futreal PA, et al. A census of human cancer genes. Nat Rev Cancer 2004; 4(3): 177–183

Garraway LA, et al. Lessons from the cancer genome. Cell 2013; 153(1): 17–37

Haber DA, Settleman J. Cancer: drivers and passengers. Nature 2007; 446(7132): 145–146

Krzywinski M, et al. Circos: an information aesthetic for comparative genomics. Genome Res 2009; 19 (9): 1639–1645

Ledford H. Big science: The cancer genome challenge. Nature 2010; 464(7291): 972–974

Lee W, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature 2010; 465(7297): 473–477

Martincorena I, et al. Somatic mutation in cancer and normal cells. Science 2015; 349(6255): 1483–1489

Polak P, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature 2015; 518(7539): 360–364

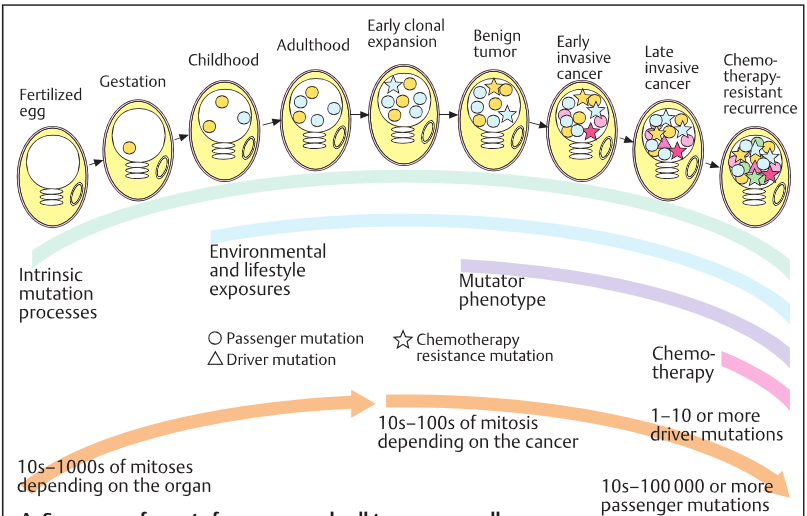
Stephens PJ, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature 2009; 462(7276): 1005–1010

Stratton MR, et al. The cancer genome. Nature 2009; 458(7239): 719–724

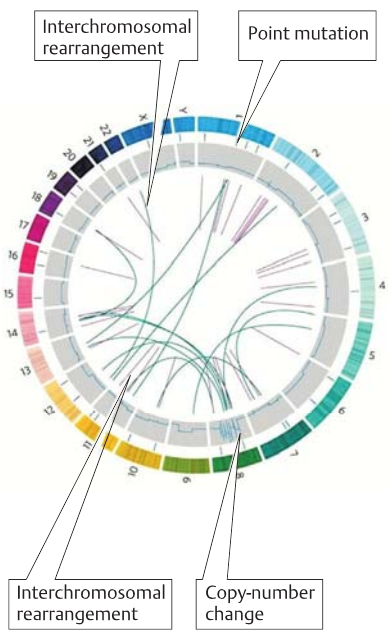
Vogelstein B, et al. Cancer genome landscapes. Science 2013; 339(6127): 1546–1558

Weinberg RA. The Biology of Cancer. 2nd ed. New York, NY: Garland Science, 2013

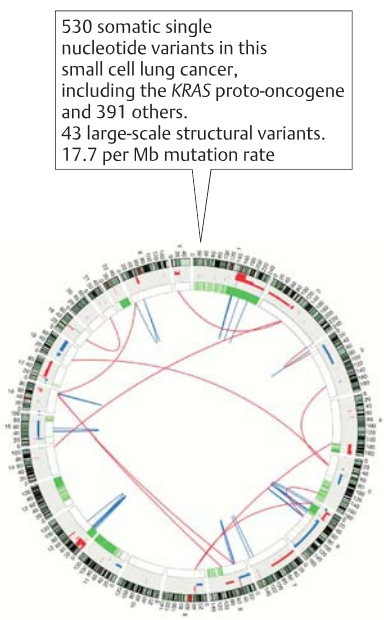
Yates LR, et al. Evolution of the cancer genome. Nat Rev Genet 2012; 13(11): 795–806



A. Sequence of events from a normal cell to a cancer cell



B. Circos plot



C. Genomic profile of an adenocarcinoma

The *TP53* Tumor Suppressor Gene

The *TP53* tumor suppressor gene (191170) is a central regulator of cell cycle control, DNA damage recognition and repair, apoptosis, and other fundamental cellular functions (called "guardian of the genome" by D.P. Lane, one of its discoverers). It derived its name from a 53-kDa nuclear phosphoprotein it encodes on the short arm of human chromosome 17 (17p13.1). The p53 protein binds to specific DNA sequences and controls the expression of different regulator genes involved in cell division and its control. Somatic mutations in the *TP53* gene occur in more than 50% of all tumors.

A. Function of the TP53

Normally, the *TP53* gene is inactive. This allows the cell to proceed through mitosis (**1**). It is activated in response to DNA damage and other cellular stress (**2**). If the subsequent DNA repair is successful, the cell may continue through mitosis. If cell cycle control is lost or repair fails, the cell undergoes apoptosis (p. 100). Mutated *TP53* results in damaged cells. This eventually leads to tumor formation. (Figure adapted from Lane, 1992.)

B. The human p53 protein

The TP53 protein is transcribed from a gene spanning 20-kb DNA with 11 exons and 2 promoters, translated from 2.8-kb mRNA. It is a transcription factor with five distinct domains that correspond to its multiple functions. In unstressed normal cells, it is kept inactive by ubiquitin ligase MDM2 (164785). The active form of the human p53 protein is a tetramer of four identical subunits. They bind to promoters of several target genes. Normally, p53 is extremely unstable because it is bound to a protein MDM2 (164785; its name is derived from mouse double minute chromosomes in mouse sarcoma). Phosphorylation at the MDM2-binding site displaces MDM2 and activates p53. TP53 has sequence-specific DNA-binding sites and interacts with several other proteins involved in DNA damage recognition and repair. Oncogenic mutations in TP53 have a dominant-negative effect. Most mutations are clustered in the conserved domains II to V: codons 129 to 146 (exon 4), 171 to 179 (exon 5), 234 to 260 (exon 7), and 270 to 287 (exon 8). Six mutations are strikingly frequent: a replacement of conserved amino acids arginine (R) in positions 175, 248, 249, 273, and 282, and glycine (G) in position 245. Mutations are missense, insertions, and deletions. The carboxyl end beyond amino acid 300 has a nonspecific DNA interaction domain and a tetramerization domain. p53 protein function is inhibited by binding to some virus proteins (papilloma virus protein E6, adenovirus protein E1b, SV40, and others). Knockout mice develop normally, but develop tumors at a high rate. Activated benzopyrene induces mutations at codons 175, 248, and 275 in cultured bronchial epithelial cells. (Figure adapted from Lodish et al., 2016.)

C. Germline mutations of *TP53*

Germline mutations lead to one of several familial cancer syndromes. One form is the autosomal dominant Li–Fraumeni syndrome (OMIM 151623) described in 1969 by Li and Fraumeni in families with diverse types of tumors. The pedigree (**1**) shows four individuals (dark symbols marked mt) with different types of tumor caused by a mutation as shown (R248W). This mutation is also present in individuals I-1 and III-5; they have not developed tumors, but are at increased risk (data from Malkin et al., 2016). The distribution of tumors (**2**) is wide, mainly involving soft-tissue sarcomas, but also early-onset breast cancer, brain cancers, cancer of the bone (osteosarcoma), and bone marrow (leukemias), and carcinoma of the lung, pancreas, and adrenal cortex. Li–Fraumeni syndrome can also be caused by mutations in the *CHEK2* gene (609265) and *LFS3* gene (609266).

Further Reading

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144(5): 646–674

Lane DP. Cancer. p53, guardian of the genome. Nature 1992; 358(6381): 15–16

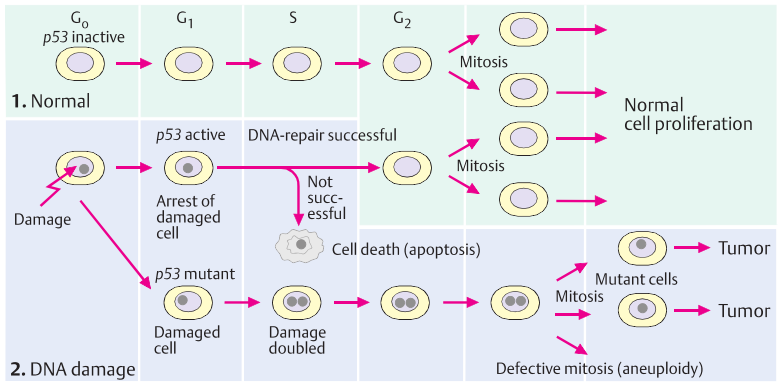
Liu Y, et al. Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. Nature 2016; 531(7595): 471–475

Lodish H et al. Molecular Cell Biology. 8th ed. WH Freeman, New York, 2016

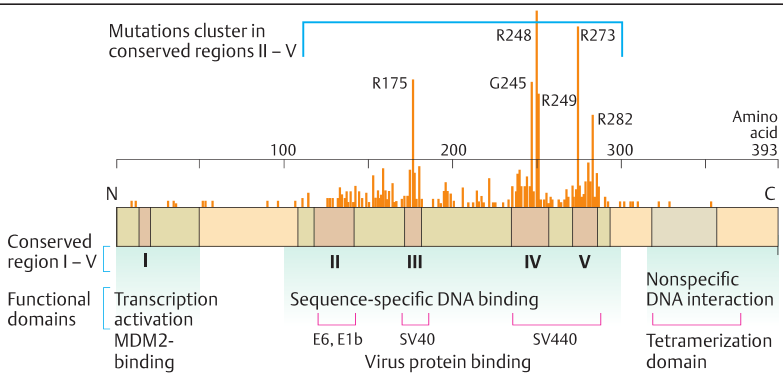
Malkin D. The Li–Fraumeni syndrome. In: Vogelstein B, et al., eds. The Genetic Basis of Human Cancer. 2nd ed. New York, NY: McGraw-Hill, 2002: 387–401

Malkin D, et al. CANCER. The cancer predisposition revolution. Science 2016; 352(6289): 1052–1053

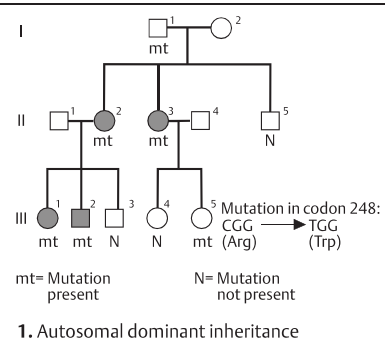
Tobias ES. The molecular biology of cancer. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medial Genetics. 6th ed. New York, NY: Elsevier, 2013



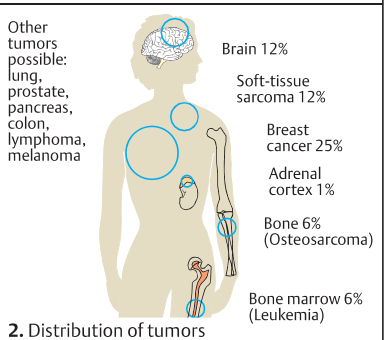
A. Model of the function of the *p53* gene



B. The human TP53 protein



1. Autosomal dominant inheritance



2. Distribution of tumors

C. Mutations of the *p53* gene in familial multiple tumors (Li-Fraumeni syndrome)

The *APC* Gene and Polyposis Coli

The *APC* gene (611731) encodes a protein with major functions in the WNT/β-catenin signal pathway (p. 190), as well as in cell migration and adhesion, apoptosis, and others. Loss of function of the *APC* gene results in adenomatous polyposis of the colon and, together with other mutations, subsequently leads to cancer progression. Mutant APC protein does not bind to β-catenin. This induces transcription of several growth-controlling genes, including the oncogene *MYC* (190080). Germline mutations in the *APC* gene are the main cause of familial adenomatous polyposis of the colon. Cancer of the colon and rectum is the second leading cause of death from cancer.

A. Familial adenomatous polyposis

Familial adenomatous polyposis coli (FAP, 175100) is an autosomal dominant hereditary disease. In late childhood and early adulthood, up to 1,000 and more polyps develop in the mucous membrane of the colon (**1**). In any polyp, an adenoma can develop (**2**). An adenoma will develop into a carcinoma (**3**). (1 and 3 kindly provided by U. Pfeifer, Bonn, Germany; 2 from Vogelstein et al., 2013.)

B. Structure and function of the *APC* gene

The *APC* gene (611731), located on human chromosome 5q22.2, has 8,538 bp in 15 exons. It encodes a protein of 2,843 amino acids with several alternatively spliced forms. Exon 15 has an exceptionally long open reading frame of 6,579 bp. Over 95% of mutations result in a nonfunctional truncated protein with variable loss of the C-terminus resulting from nonsense mutations (40%), deletions (41%), insertions (12%), and splice site mutations (7%). The site of the mutation influences the phenotypic manifestations.

C. Diagnosis in FAP

This example illustrates the formerly used indirect DNA analysis by haplotype analysis using polymorphic DNA marker loci flanking the APC locus (**1**) rather than directly analysis used now. Since two affected individuals (I-1 and II-3) share haplotype 6 to 8 at loci D5S28 and D5S346, this haplotype must carry the mutation. Individual II-4 also inherited this haplotype and must be considered at risk of developing FAP. The protein truncation test (**2**) detects the abnormal protein that migrates faster than the normal protein because of its smaller size. (Data kindly supplied by W. Friedl, Bonn, Germany.)

D. Tumor progression in stages

Tumor formation progresses through several stages. After the first mutation has occurred in the *APC* gene, a second event may be loss of the normal allele (LOH) or a mutation in the second allele. Additional mutations in other growth-controlling genes are associated with increasing loss of differentiated cells in the gastrointestinal mucosa, formation of adenoma, polyps, and tumor development. Approximately half of colorectal cancers have RAS mutations (603384). Other genes involved include *DCC* (120470), *TP53* (191170), *SMAD4* (600993), *SMAD2* (601366), and others. (Figure adapted from Fearon and Vogelstein, 1990.)

Hereditary nonpolyposis colorectal cancer (HNPCC, 120435) affects approximately 1 in 200 to 1,000 individuals (3% of all colorectal cancers). It results from germline mutation in one of the DNA-mismatch repair genes *hMSH1*, *hMLH2*, *hPMS1*, or *hPMS2*, or a related gene. Microsatellite instability is an important feature of HNPCC.

Further Reading

Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990; 61(5): 759–767

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144(5): 646–674

Li J, et al. Point mutations in exon 1B of APC reveal gastric adenocarcinoma and proximal polyposis of the stomach as a familial adenomatous polyposis variant. Am J Hum Genet 2016; 98(5): 830–842

Morin PJ, et al. Cancer genetics. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 663–672

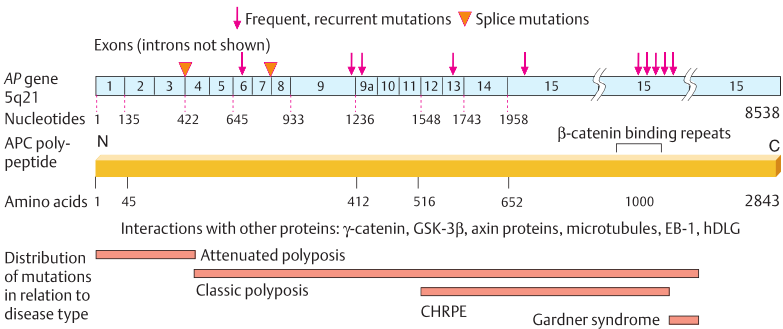
Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012; 487(7407): 330–337

Vogelstein B, et al. Cancer genome landscapes. Science 2013; 339(6127): 1546–1558

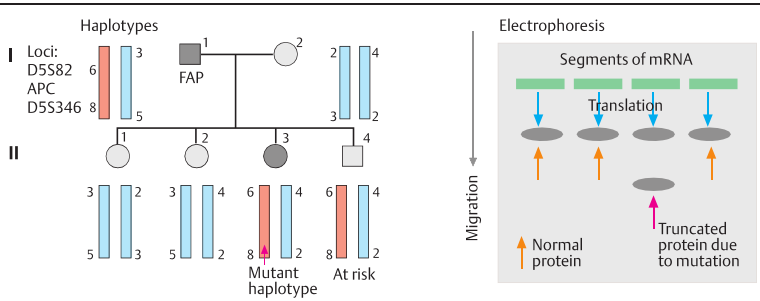
Weinberg RA. The Biology of Cancer. 2nd ed. New York, NY: Garland Science, 2013

1. **2.** **3.**

A. Polyposis coli and colon carcinoma



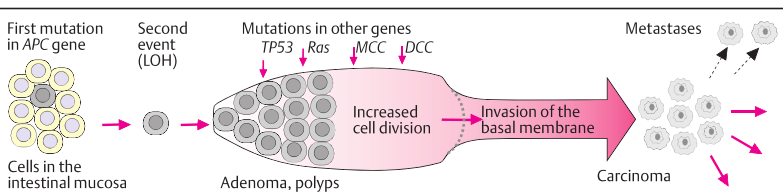
B. Structure and function of the *APC* gene



1. Haplotype analysis

2. Protein truncation test

C. Indirect and direct DNA diagnosis in FAP



D. Tumor progression in stages

Breast and Ovarian Cancer Susceptibility Genes

Two major genes and more than 20 others confer susceptibility to breast and ovarian cancer when mutated. The identification of two breast cancer–related genes *BRCA1* and *BRCA2* in 1994 was a major discovery. Both genes encode multifunctional proteins with important roles in genomic stability and DNA repair (see p. 80). They interact with other proteins and participate in cell cycle control (see p. 90). In addition to the usual somatic mutations, germline mutations are the basis for familial occurrence in approximately 10 to 15%. The causative role of a particular sequence change may be difficult to assess in a given individual. In addition, polymorphic sequence variants are common. Mutations in approximately 90 other cancer genes are involved, in particular in DNA double-strand break repair pathways such as *ATM*, *ATR*, *CHEK2*, *RAD51C*, *RAD50*, *ATM*, *BRIP1*, and *PALB2* (Nik-Zainal et al., 2016).

A. The breast cancer susceptibility gene BRCA1

The *BRCA1* gene (113705, 604370) on chromosome 17 at q21.31 accounts for 20 to 30% of inherited, autosomal dominant forms of breast cancer. In 24 exons spanning 8 kb of genomic DNA, it encodes an mRNA transcript of 7.8 kb. Somatic mutations in breast tissue and germline mutations observed in unrelated patients are evenly distributed throughout the gene (red arrows). A deletion of an adenine (A) and a guanine (G) in nucleotide position 185 (185delAG) and an insertion of a cytosine in position 5382 (5382insC) are the most frequent mutations, each accounting for approximately 10% of mutations. These mutations are particularly frequent in Ashkenazi Jewish population.

The protein with1,863 amino acids has distinct functional domains: (BRCA1-associated RING domain 1), three protein-binding domains that interact with the TP53 protein, the DNA recombinatuion protein RAD51 (a human homolog of the bacterial RecA protein), and an RNA helicase. RAD50 (604040) and RAD51 (179617) proteins are essential for repair of double-strand DNA breaks. The C-terminus contains a region involved in transcriptional activation and DNA repair. Two nuclear localization signals (NLS) are present at amino acid positions 500 to 508 and 609 to 615.

B. The breast cancer susceptibility gene *BRCA2*

Mutations in the *BRCA2* gene (600185) at 13q12.3 occur throughout the gene. A deletion of thymine at nucleotide position 6174 (6174delT) is relatively frequent (1%) in the Ashkenazi Jewish population. The BRCA2 protein also has distinct functional domains. A large central domain consists of eight copies of a repeat of 30 to 80 amino acids. Four of these interact with the RAD51 protein. *BRCA2* is identical to *FANCD1*, a gene in the Fanconi anemia pathway (p. 330). (Figures redrawn from Couch and Weber, 2002, and Welcsh et al., 2000.)

C. BRCA1-mediated effect on TP53

BRCA1 functions as a coactivator of transcription of several genes. Particularly important is the interaction with the TP53 protein (p. 318) in inducing apoptosis in damaged cells. Eukaryotic cells have two pathways to deal with double-stranded DNA breaks: nonhomologous end-joining and homologous recombination (HR; the mechanism used for recombination during meiosis, see p. 94). BRCA1 is connected to HR. (Figure redrawn from Hohenstein and Giles, 2003.)

Further Reading

Couch FJ, et al. Breast cancer. In: Vogelstein B, et al., eds. The Genetic Basis of Human Cancer. 2nd ed. New York, NY: McGraw-Hill, 2002: 549–581

Hohenstein P, Giles RH. BRCA1: a scaffold for p53 response? Trends Genet 2003; 19(9): 489–494

Lawrence MS, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014; 505(7484): 495–501

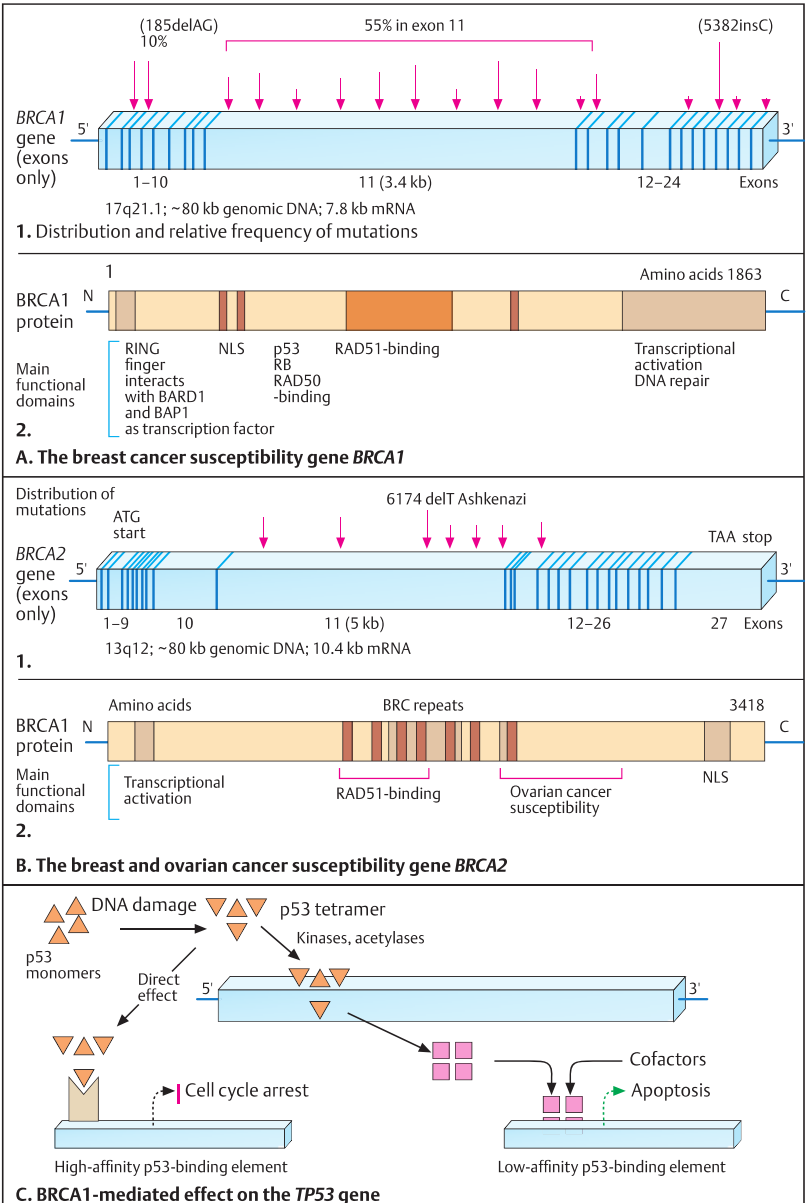
Meindl A, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet 2010; 42(5): 410–414

Nik-Zainal S, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016; 534(7605): 47–54

Ripperger T, et al. Breast cancer susceptibility: current knowledge and implications for genetic counselling. Eur J Hum Genet 2009; 17(6): 722–731

Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012; 490(7418): 61–70

Welcsh PL, et al. Insights into the functions of BRCA1 and BRCA2. Trends Genet 2000; 16(2): 69–74


(185delAG)
10%
55% in exon 11
(5382insC)
BRCA1
gene
(exons
only)
5'
3'
1–10
11 (3.4 kb)
12–24
Exons
17q21.1; ~80 kb genomic DNA; 7.8 kb mRNA
1. Distribution and relative frequency of mutations
1
Amino acids 1863
BRCA1
protein
N
C
Main
functional
domains
RING
finger
interacts
with BARD1
and BAP1
as transcription factor
NLS
p53
RB
RAD50
-binding
RAD51-binding
Transcriptional
activation
DNA repair
2.
A. The breast cancer susceptibility gene BRCA1
Distribution of
mutations
ATG
start
6174 delT Ashkenazi
TAA stop
BRCA2
gene
(exons
only)
5'
3'
1–9
10
11 (5 kb)
12–26
27
Exons
13q12; ~80 kb genomic DNA; 10.4 kb mRNA
1.
Amino acids
BRC repeats
3418
BRCA1
protein
N
C
Main
functional
domains
Transcriptional
activation
RAD51-binding
Ovarian cancer
susceptibility
NLS
2.
B. The breast and ovarian cancer susceptibility gene BRCA2
DNA damage
p53 tetramer
Kinases, acetylases
p53
monomers
Direct
effect
5'
3'
Cofactors
Cell cycle arrest
Apoptosis
High-affinity p53-binding element
Low-affinity p53-binding element
C. BRCA1-mediated effect on the TP53 gene

Oncogenic Chromosome Translocations

Genes with a control function in cell division are a common cause of leukemias and lymphomas. A reciprocal translocation involving two such genes leads to a fusion of parts of the genes. As a result, a new protein consisting of fragments of two proteins with altered disease-causing functions is formed. A common example is chronic myeloid leukemia (CML). Several other oncogenic translocations are listed in the Appendix in Table 13, p. 413.

A. Principal features of CML

CML (608232) usually begins in adults with a greatly increased number of myelocytes (white blood cells, stained blue) in the bone marrow and peripheral blood (**1**). Its hallmark is a characteristic chromosome abnormality in myeloid cells, the Philadelphia chromosome (Ph^1), a reciprocal translocation between a chromosome 9 and a chromosome 22 (**2**, arrows). (Images by Dr. A. Schneider, Essen.)

B. Ph^1 translocation

The Ph^1 chromosome was recognized as the first specific chromosome aberration associated with a leukemia by Nowell and Hungerford in Philadelphia in 1960 and 1973 by Janet D. Rowley in Chicago. The breakpoints of the reciprocal translocation (**1,** small arrows) disrupt two genes on chromosomes 22 and 9 (**2**). Approximately half of the long arm of a chromosome 22 is translocated to the long arm of a chromosome 9, and a very small part of the distal long arm of a chromosome 9 is translocated to a chromosome 22, but this is not visible by light microscopy.

C. Fusion of two genes: *BCR* and *ABL*

The breakpoints of the translocation are in the *BCR* gene 151410) on chromosome 22q11.21 and the *ABL* gene (189980) on chromosome 9q34.12. As a result, the chromosome parts carrying these two genes are fused (151410). In the *BCR* gene, the breakpoints cluster to a small region of 5.8 kb (breakpoint cluster region, BCR) in exons 10 to 12 of this gene, and to a larger region of 180 kb between exons 1a and 1b in the *ABL* gene. In contrast, in a different type of leukemia, acute Ph^1-positive leukemia (e.g., acute lymphocytic leukemia, ALL) the breakpoints are localized in the 5' direction in exons 1 or 2.

D. The BCR/ABL fusion protein

The *ABL* gene encodes a tyrosine kinase with functions in cellular growth controls. It is transcribed into two alternative mRNA transcripts of 7 kb (exon 1b, 2–11) and 6 kb (exon 1a, 2–11) length (**1**), and translated into a 145-kDa protein called $p145^{abl}$ (**2**). The fused genes are transcribed into an mRNA transcript of 8.5 kb (**3**). The fusion protein (**4**) is a 210-kDa protein ($p210^{bcr/abl}$) that has an inappropriately active ABL kinase domain. As a result, hematopoietic cells in the bone marrow may proliferate excessively requiring therapy. A small molecule, STI571 (**5**), a tyrosine inhibitor, precisely fits into the fusion protein and abolishes its abnormal function (**6**). It is used for therapy in CML under the name Gleevec along with other small molecules serving as tyrosine inhibitors exist. Resistance to these compounds may develop as a result of new mutations in the *BCR/ABL* fusion gene. (6, adapted from Schindler et al., 2000.)

Further Reading

Druker BJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 2001; 344(14): 1031–1037

Melo JV, et al. Chronic myeloid leukaemia as a model of disease evolution in human cancer. Nat Rev Cancer 2007; 7(6): 441–453

Nowell PC, et al. A minute chromosome in human chronic granulocytic leukemia. Science 1960; 132: 1497

Nowell PC. Discovery of the Philadelphia chromosome: a personal perspective. J Clin Invest 2007; 117(8): 2033–2035

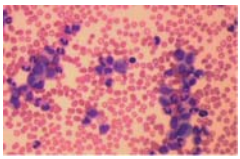
Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 1973; 243(5405): 290–293

Schindler T, et al. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science 2000; 289(5486): 1938–1942

CML Support. Available at: www.cmlsupport.org.uk/. Accessed March 20, 2017

Wetzler M, et al. Acute and chronic myeloid leukemia. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 905–918

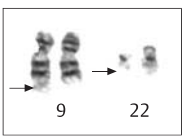
Zhang Y, et al. Leukemias, lymphomas, and other related disorders. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013



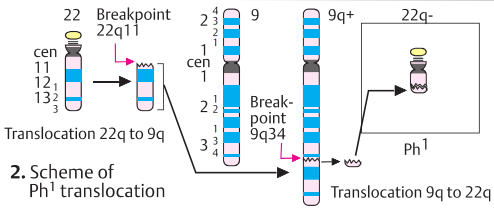
1. Accumulation of white blood cells (blue)

2. The Philadelphia translocation 9q;22q

A. Main features

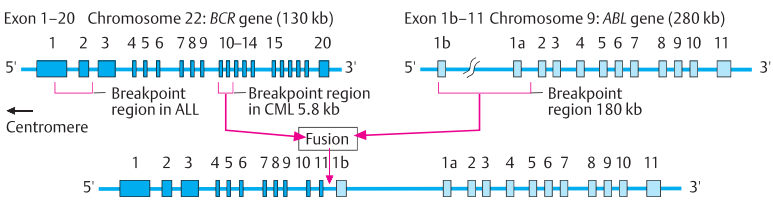


1. Ph^1 translocation

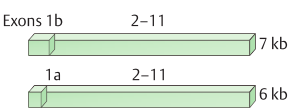


2. Scheme of Ph^1 translocation

B. Ph^1 translocation [t(9;22) (q34;q11)]

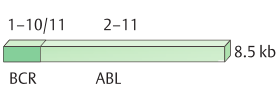


C. Ph^1 translocation causes fusion of two genes



1. Normal ABL mRNA

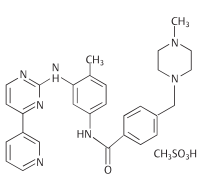
2. Normal ABL protein (145 kDa)



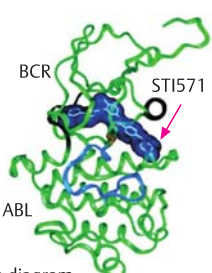
3. Abnormal BCR/ABL fusion

D. The BCR/ABL fusion protein

4. Abnormal fusion protein (210 kDa)



5. Chemical structure of STI571 (Gleevec)



6. Ribbon diagram

Retinoblastoma

Retinoblastoma (180200) is the most frequent malignant tumor of the eye in infancy and early childhood, with an incidence of approximately 1 in 15,000 to 25,000 live births. It results from loss of function of both alleles of the retinoblastoma gene *RB1* (614041) in undifferentiated cells (retinoblasts) in the retina. *RB1* is the first tumor suppressor gene, identified in 1986. Thus, two inactivating events are required for tumor initiation, as predicted by A. Knudson in 1971 (two-hit hypothesis). The first mutation predisposes the cell to develop a tumor, and the second initiates tumor formation.

A. Phenotype

Retinoblastoma (Rb) may occur in one or both eyes. An early sign is a white shimmer, called "cat's eye" (**1**), and/or rapidly developing strabismus. One or several tumors (unifocal or multifocal, respectively) may be present in the retina of an affected eye (**2**) and progress rapidly (**3**). Early diagnosis and therapy are essential. In some families, carriers of an oncogenic mutation do not develop tumors (nonpenetrance). This low-penetrance phenotype is associated with specific *RB1* mutations. Milder phenotypic expression is also observed when the mutation is present in only a proportion of germ cells (mutational mosaicism). (Images courtesy of D. Lohmann and the late W. Höpping, Essen.)

B. Retinoblastoma locus

The *RB1* locus at 13q14.2 was first identified by microscopically visible interstitial deletions. Approximately 60% of patients have somatic mutations (nonhereditary Rb) and usually develop unilateral originating from one site (unifocal Rb). Approximately 40% of patients are heterozygous for an *RB1* mutation that is either transmitted from one parent as an autosomal dominant trait (10–15%) or is the result of a new mutation, usually in the paternal allele (approximately 10:1).

C. Retinoblastoma gene *RB1* and its protein

The *RB1* gene is organized into 27 exons of different sizes (31–1,889 bp) spanning 200-kb genomic DNA (**1**). It is ubiquitously expressed and transcribed into mRNA of 4.7 kb (**2**). The gene product (pRB protein), a 100-kDa phosphoprotein with 928 amino acids (**3**), has important functions in the regulation of the cell cycle (p. 90). It is activated by phosphorylation at approximately 12 distinct serine and threonine residues (P) during cell cycle progression from G_0 to G_1. Three functional domains, A, B, and C, bind in a cell-cycle-dependent fashion to transcription factors of other genes. A nuclear localization signal (NLS) is located at the C-terminus. Kanber et al. in 2009 showed that the *RB1* gene is imprinted.

D. Diagnostic principles

In approximately 3–5% of patients, an interstitial deletion of 13q14 or a larger 13q area is visible by chromosomal analysis (**1**), usually associated with signs of developmental delay. Formerly indirect DNA diagnosis using intragenic DNA markers is now replaced by direct gene analysis. In this example, intragenic haplotype analysis (**2**) of an affected girl (II-1) reveals that in tumor cells the paternal haplotype (a -3332) only was present due to loss of heterozygosity, LOH) and carried the mutation. An example of direct DNA analysis in familial Rb (**3**) prior to sequencing analysis available today demonstrates a germline mutation in blood cells (**4**). Two affected individuals (I-2 and II-2) carry a disease-related mutation (a cytosine to thymine in codon 575). This changed a CAA glutamine (Q) to a TAA stop codon (X). (Photographs courtesy of W. Höpping and D. Lohmann, Essen, Germany.)

Further Reading

Dommering CJ, et al. RB1 mutation spectrum in a comprehensive nationwide cohort of retinoblastoma patients. J Med Genet 2014; 51(6): 366–374

Eloy P, et al. A parent-of-origin effect impacts the phenotype in low penetrance retinoblastoma families. PLoS Genet 2016: e1005888

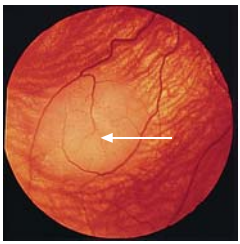
Kanber D, et al. The human retinoblastoma gene is imprinted. PLoS Genet 2009; 5(12): e1000790

Lohmann DR, Gallie BL. Retinoblastoma: revisiting the model prototype of inherited cancer. Am J Med Genet 2004; 129C: 23–28

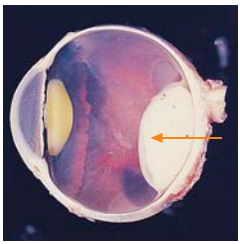
Retinoblastoma Eye Cancer Network. Available at: http://www.eyecancer.com/ (Accessed 4 September 2017).

Zhang J, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature 2012; 481(7381): 329–334

1. So-called cat's eye

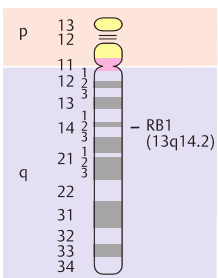


2. Tumor in the retina



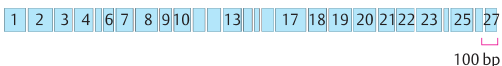
3. Large tumor in the eye

A. Phenotype

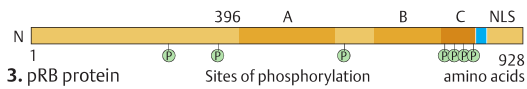


B. Retinoblastoma locus on chromosome 13

1. Exon/Intron structure

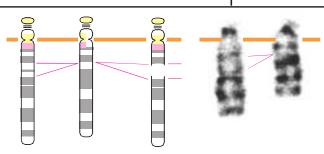


2. Coding regions (27 exons)

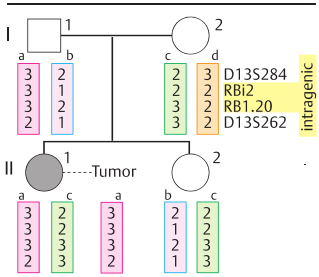


3. pRB protein Sites of phosphorylation

C. Retinoblastoma protein (pRB1) with functional domains A–C

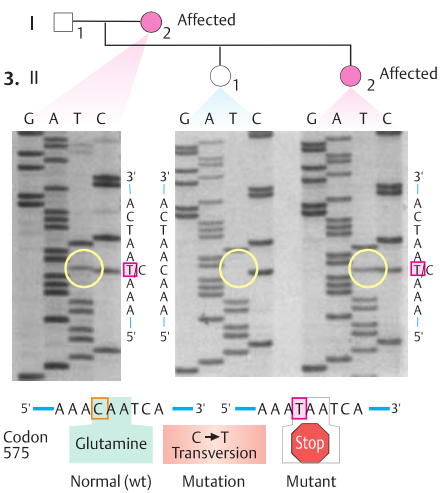


1. Interstitial deletion



2. Haplotype analysis

D. Diagnostic principle



3. **4.** Sequence analysis

Neurofibromatosis

The neurofibromatoses are a group of clinically and genetically different autosomal dominant diseases that predispose to benign and malignant tumors of the nervous system. Most common among others are neurofibromatosis 1 (NF1,162200) and neurofibromatosis 2 (NF2,607379).

A. Main manifestations of neurofibromatosis 1

NF1, also called von Recklinghausen disease, affects 1 in 3,000 individuals with highly variable manifestations. Characteristic are Lisch nodules of the iris (**1**) in more than 90% of patients, café-au-lait spots (**2**) (more than five spots of more than 2-cm diameter are considered diagnostic) in more than 95% of patients, and multiple neurofibromas (**3**) in more than 90% of patients, usually apparent between the ages of 4 and 15 years. Approximately 2 to 3% of patients develop a neurofibrosarcoma or other malignancies. The NF1 phenotype may overlap with Noonan syndrome (601321).

B. The *NF1* gene

The neurofibromin gene (NF1, 613113) is large, spanning 350-kb genomic DNA, and organized into at least 59 exons. It is a tumor suppressor gene localized on human chromosome 17 at q11.2. The gene was isolated in 1990 from a 600-kb *Nru*I restriction fragment by positional cloning. Two patients with a translocation involving the long arm of chromosome 17 with breakpoints at 17q11.2 and CpG islands provided anchor points. Three unrelated genes, *OMGP* (164345), *EVI2B* (158381), and *EVI2A* (158380), are located within an intron of *NF1*, but in the opposite DNA strand. The types of *NF1* mutations are deletions, insertions, base substitutions, and splice mutations. *NF1* has one of the highest mutation rates, with a new mutation in approximately 50% of patients. Numerous *NF1* pseudogenes are present in the human genome.

C. *NF1* gene product, neurofibromin-1

The gene product, neurofibromin-1, is a multidomain cytoplasmic protein, which regulates several intracellular signal processes, including the RAS/MAPK pathway (p. 252). The human neurofibromin is a protein of 2,818 amino acids. It is transcribed from multiple alternatively spliced transcripts and expressed in many tissues. A region of GAP homology with a gene product in yeast (*Saccharomyces cerevisiae*), IRA1 (inhibitor of RAS mutants), is located between amino acids 840 and 1,200. (Figure adapted from Xu et al., 1990.)

D. Neurofibromatosis 2 gene, *NF2*

Neurofibromatosis 2 (101000) is characterized by multiple tumors of the eighth cranial nerve, mainly meningioma, often bilateral, with an incidence of 1 in 25,000. The *NF2* gene (607379), localized on human chromosome 22 at q12.2, is organized into 17 constitutive exons and one alternatively spliced exon, spanning approximately 110 kb of genomic DNA. The widely expressed gene product, neurofibromin-2 (also called schwannomin or merlin), regulates cell proliferation, cell–cell adhesion, transmembrane signaling, and other functions. It is a member of the band 4.1 cytoskeleton-associated protein superfamily (see p. 334) called the ERM family (including ezrin, radixin, moesin, and several protein tyrosine phosphatases). The gene was identified within a cosmid contig of YAC (yeast artificial chromosome) clones. Two deletions (Del 1 and Del 2) in unrelated patients were useful in the process of finding the gene. Mutations can be detected in more than 50% of patients (large deletions, including the entire gene or several exons, and small deletions are frequent). (Figure adapted from Rouleau et al., 1993.)

Further Reading

Asthagiri AR, et al. Neurofibromatosis type 2. Lancet 2009; 373(9679): 1974–1986

Crowe FW, et al. Clincical, Pathological, and Genetic Study of Multiple Neurofibromatosis. Springfield, IL: Charles C Thomas, 1956

Huson SM, et al. The Neurofibromatoses. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Jett K, et al. Clinical and genetic aspects of neurofibromatosis 1. Genet Med 2010; 12(1): 1–11

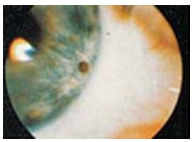
Riccardi VM, et al. Neurofibromatosis. Phenotype, Natural History and Pathogenesis. 2nd ed. Baltimore, MD: Johns Hopkins University Press, 1992

Rouleau GA et al. Alteration in a new gene encoding a putative membrane-organizing protein causes neurofibromatosis type 2. Nature 1993; 363 (6329): 515–521

Xu GF, et al. The neurofibromatosis type 1 gene encodes a protein related to GAP. Cell 1990; 62(3): 599–608

Neurofibromatosis 1 (NF1)
(von Recklinghausen disease)

Autosomal dominant
Frequency 1 in 3000
Gene locus on 17q11.2
Café-au-lait spots
Lisch nodules in the iris
Multiple neurofibromas
Skeletal anomalies
Predisposition to tumors of the nervous system
50% new mutations

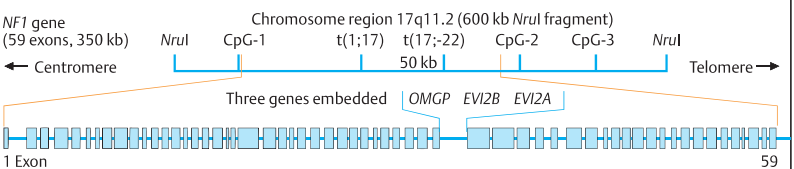


1. Lisch nodule

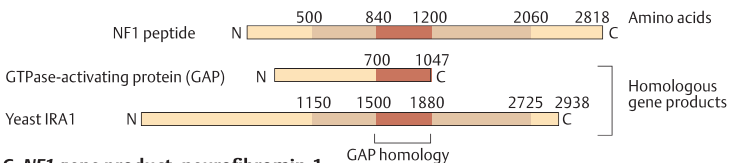
2. Café-au-lait spot

3. Neurofibromas

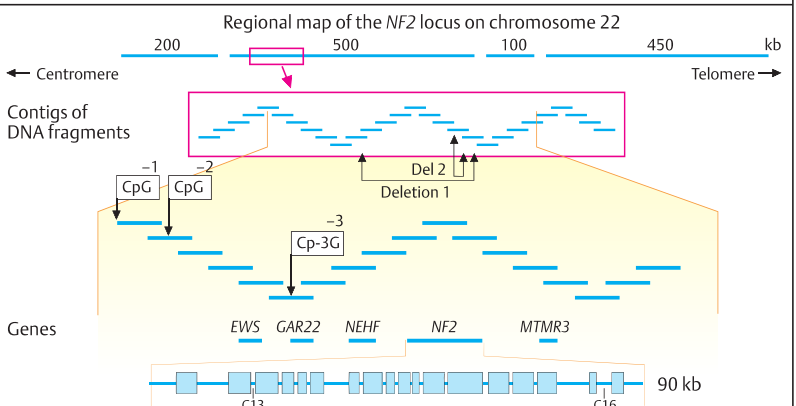
A. Main manifestations of neurofibromatosis 1



B. The *NF1* gene



C. *NF1* gene product, neurofibromin-1



D. Neurofibromatosis gene *NF2* on chromosome 22q12.1

Genomic Instability Diseases

Genomic instability diseases are a group of different disorders that share an increased risk to malignant tumors and a cellular phenotype consistent with genomic instability. Mutations in their genes result in different defects of DNA damage recognition and repair. Ataxia-telangiectasia, Fanconi anemia, and Bloom syndrome are important examples. Different patterns of chromosomal breaks and rearrangements are visible by light microscopy of metaphase cells.

A. Ataxia-telangiectasia (AT)

AT (208900) is a variable autosomal recessive disease resulting from mutations in the *ATM* gene at (607585) at 11q22.3 with an incidence of 1:40,000. The main manifestations, which develop in early childhood, are cerebellar ataxia, immune defects, and characteristic telangiectasias of the conjunctivae (**1**). Affected individuals are highly sensitive to irradiation and are prone to lymphomas and leukemias. The *ATM* gene has 66 exons spanning 150-kb genomic DNA. The ATM protein consists of 3,056 amino acids and is a member of the phosphatidylinositol 3-protein kinase family that respond to DNA damage (PIK3, 601232). ATM is activated in response to double-strand DNA breaks. It has a central role in a network of proteins that regulate cellular responses to DNA damage and recombination (**2**). Heterozygotes also are predisposed to tumors. Mutations in a functionally related gene result in the Nijmegen breakage syndrome (NBS1,251260).

B. Fanconi anemia (FA)

FA (227650) is a genetically highly heterogeneous group of 19 subtypes (see Appendix, Table 14, p. 413). It manifests in early childhood as pancytopenia, growth deficiency (**1**), hypoplastic radius often with hypoplastic or absent thumbs (**2**), and other malformations. The subtypes are assigned to 19 complementation groups named FANCA, B, C, D2, E, F, G, I, J, L, M, N, O, P, Q, R, T, U, and V (see phenotypic series in OMIM 227650). Seven proteins FANCA, -B, -C, -E, -F, -G, and L form a nuclear core complex as part of an FA/BRCA pathway. They act in response to DNA damage and in cooperation with several other proteins (**3**, TP53, ATM, BRCA1, BRCA2, NBS1, BLM, RAD50/51). FANCB is the only X-chromosomal (Xp22.2); all others are autosomal. FANCD2 with a central role is activated by addition on an ubiquitin (Ub) in response to DNA damage. The most prevalent mutation is FANCA, in approximately 65% of patients. FA cells are hypersensitive to DNA cross-linking agents, such as diepoxybutane (DEB). *FANCD1* and BRCA2 are identical. (3, adapted from Schindler and Hoehn, 2007.)

C. Bloom syndrome (BLM)

BLM (210900) is a prenatal and postnatal growth deficiency disorder (birth weight 2,000 g, birth length 40 cm, adult height approximately 150 cm) with a distinct phenotype (**1**): narrow face, sunlight-induced facial erythema, variable immune deficiency, and a greatly increased risk of different malignancies (approximately one in five patients). Chemotherapy is very poorly tolerated. The diagnostic hallmark is a 10-fold increase in the spontaneous rate of sister chromatid exchanges (**2**, **3**). BLM results from autosomal recessive mutations in the *RECQL3* gene (604610) at 15q26.1, encoding a member of the RecQ family of DNA helicases. The mutations are mainly protein-truncating nonsense mutations, and they are distributed fairly evenly along the gene (**4**), but some missense mutations occur. A founder mutation occurs in populations of Ashkenazi Jewish origin: a 6-bp deletion/7-bp insertion at nucleotide 2281. (Figure in 4, courtesy of N.A. Ellis and J. German, New York.)

Further Reading

NIH Genetics Home Reference. Ataxia-telangiectasia. Available at: https://ghr.nlm.nih.gov/condition/ataxia-telangiectasia. Accessed March 22, 2017

Auerbach AD. Fanconi anemia and its diagnosis. Mutat Res 2009; 668(1–2): 4–10

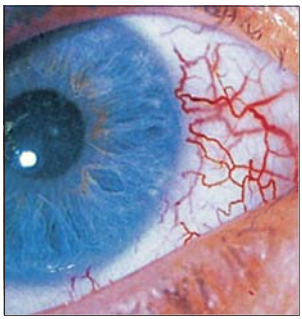
Deakyne JS, et al. Fanconi anemia: at the crossroads of DNA repair. Biochemistry (Mosc) 2011; 76(1): 36–48

Huang Y, et al. Modularized functions of the Fanconi anemia core complex. Cell Reports 2014; 7(6): 1849–1857

Kim H, et al. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. Genes Dev 2012; 26(13): 1393–1408

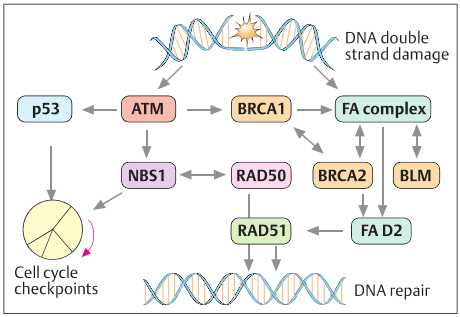
Sanz MM, et al. Bloom's syndrome. March 22, 2006 [updated 2016]. In: GeneReviews®. Available at: www.ncbi.nlm.nih.gov/books/NBK1116/. Accessed March 22, 2017

Schindler D, Hoehn H, eds. Fanconi Anemia. A Paradigmatic Disease for the Understanding of Cancer and Aging. Basel: Karger, 2007



1. Telangiectasias of the conjunctiva

A. Ataxia-telangiectasia (AT)

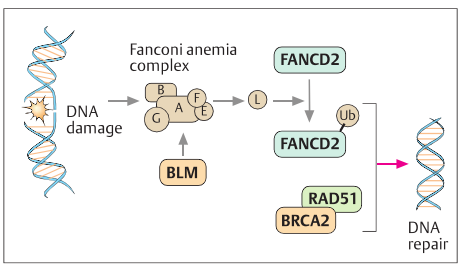


2. Relation of ATM to other proteins maintaining genomic stability

2. Thumb hypoplasia

1. Phenotype

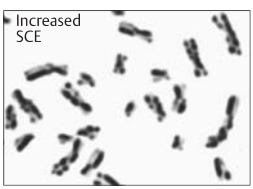
B. Fanconi anemia (FA)



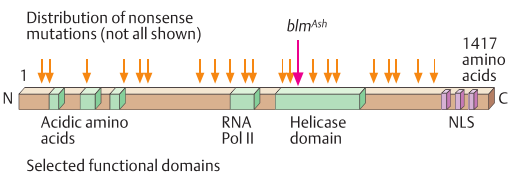
3. Fanconi anemia-associated proteins

a b c

1. Phenotype

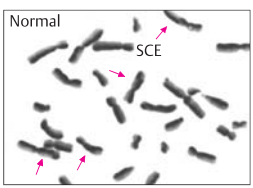


2. Bloom syndrome metaphase



4. Diagram of the Bloom protein and distribution of mutations

C. Bloom syndrome



3. Normal metaphase

DNA Excision Repair Disorders

Nucleotide excision repair (NER) is a major form of DNA repair (see p. 80). Three main examples of this genetically heterogeneous group of disorders are presented here: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy.

A. Xeroderma pigmentosum (XP)

The leading signs of XP (278700) are dry and pigmented skin in ultraviolet light (UV)-exposed regions of the skin (**1**,**2**), from which the name was derived, when it was first described by F. von Hebra and M. Kaposi in 1874. Unexposed areas are not affected. Thus, it is important to protect patients from UV light. An especially important feature is the tendency for multiple skin tumors to develop in the exposed areas (**3**). These may occur as early as childhood or early adolescence. The types of tumors are the same as those occurring in healthy individuals after prolonged UV exposure. The basic defect is DNA excision repair deficiency, as first shown by Cleaver in 1968 due to mutations in the *XPA* gene (611153). The causes are mainly autosomal recessive mutations in genes at several loci encoding one of the several excision repair proteins: XPA, -B, -C, -D, -E, -F, -G, or XPV. These genes are highly conserved in bacteria, yeast, and mammals.

B. Cellular phenotype of XP

The UV sensitivity is evident in cultured fibroblast cells. When exposed to UV light, XP cells show a distinct dose-dependent decrease in survival rate compared with normal cells (**1**). When cultured in the presence of ^{3}H-thymidine and exposed to UV light, an autoradiograph (a film that, placed on the cells) develops black dots at sites where radiolabeled thymine has been incorporated during DNA synthesis. XP cells show few dots over the nucleus of repair-deficient XP cells, here XP-D and XP-A (**2**). However, if XP cells with different defects are fused they can correct each other (as in the heterokaryon XPA/XPD). This established the existence of different types of XP, assigned to complementation groups based on their ability to complement (correct) each other. (Photograph courtesy of Bootsma and Hoeijmakers, 1991.)

C. Cockayne syndrome (CS)

The leading signs of CS (216400) are delayed development with cachectic, progeroid appearance (1), cutaneous photosensitivity and atrophy (2), and manifestations in other organ systems. Two types of CS can be distinguished, type A and B, caused by different genes at 5q12.1 (609412) and 10q11.23 (133540), respectively. In addition CS overlaps with XPG (XPG/CS (278780/133530). (Images obtained from Google in 2012.)

D. Trichothiodystrophy

This heterogeneous group consists of six genetically different members (601675) at five autosomal and one X-chromosomal loci (Xq24, 300953) with photosensitive and a nonphotosensitive forms. Ichthyotic skin (**1**) associated with developmental delay and sulfur-deficient brittle hair (**2**) and nails are the phenotypic hallmark of this group. Other abnormal hair structures are apparent by different methods of light microscopy (**3–5**), such as trichoschisis and trichorrhexis nodosalike fractures. (Figure modified from Liang et al., 2006.)

Further Reading

Boulton SJ. DNA repair: Decision at the break point. Nature 2010; 465(7296): 301–302

Bootsma D, Hoeijmakers JH. The genetic basis of xeroderma pigmentosum. Ann Genet 1991; 34(3–4): 143–150

Cleaver JE. Defective repair replication of DNA in xeroderma pigmentosum. Nature 1968; 218(5142): 652–656

Cleaver JE, et al. A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. Hum Mutat 1999; 14(1): 9–22

Cleaver JE, et al. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. Nat Rev Genet 2009; 10(11): 756–768

de Boer J, et al. Nucleotide excision repair and human syndromes. Carcinogenesis 2000; 21(3): 453–460

Hashimoto S, et al. Trichothiodystrophy view from the molecular basis of DNA repair/transcription factor TFIIH. Hum Mol Genet 2009; 18(R2): R224–R230

Liang C, et al. Structural and molecular hair abnormalities in trichothiodystrophy. J Invest Dermatol 2006; 126(10): 2210–2216

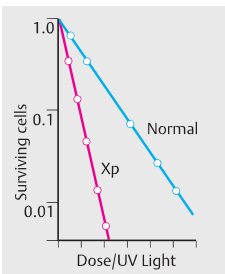
Plon P. Disorders of DNA Repair and Metabolism. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

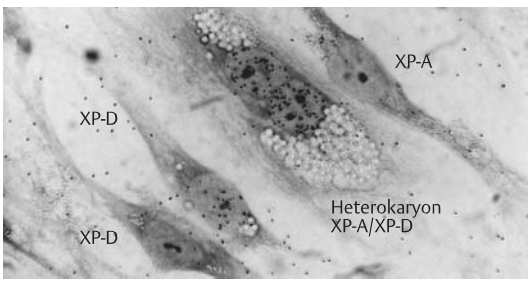
1.

2.

3.

A. Xeroderma pigmentosum: Clinical phenotype





1. UV sensitivity **2.** Complementation following fusion of XP-A and XP-D cells

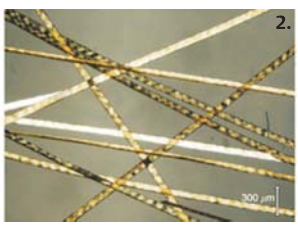
B. Xeroderma pigmentosum: Cellular phenotype

1.

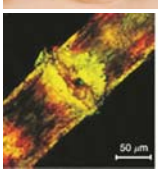
2.

C. Cockayne syndrome

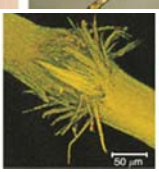
1.



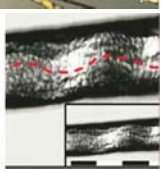
2.



3.



4.



5.

D. Trichothiodystrophy

Cytoskeletal Proteins in Erythrocytes

The cytoskeleton is an intracellular system of proteins with a fibrillar structure. The three main components are microfilaments (7.9-nm diameter), intermediate filaments (10-nm diameter), and microtubule (24-nm diameter). They consist of polymers of small subunits. Microfilaments and membrane-binding proteins serve as a skeleton of the cell under the plasma membrane. A medium-sized protein of 375 amino acids is *actin*. It is the main cytoskeleton protein (0.5×10^9 molecules) and accounts for 1 to 5% of all cellular proteins; in muscle cells, it accounts for 10%. The actin cytoskeleton is arranged in bundles and networks that are connected to each other and the cell wall. Actins form an evolutionarily conserved gene family. Erythrocytes have to meet extreme requirements: they traverse small capillaries with diameters less than that of the erythrocytes themselves approximately a half million times during a 4-month lifespan.

A. Erythrocytes

A normal erythrocyte is maintained in a characteristic biconcave discoid form by its cytoskeletal proteins. This renders the erythrocyte flexible and allows its passage through narrow capillaries. Genetic defects in different cytoskeletal proteins lead to characteristic erythrocyte deformations as shown. The various forms are the result of defects of different cytoskeleton proteins (see **D**). (Figures based on scanning electron micrographs from Davies and Lux, 1989.)

B. Proteins in the erythrocyte membrane

Spectrin is a 200-nm long, rod-shaped protein that runs parallel to the erythrocyte plasma membrane. The characteristic shape of an erythrocyte is maintained by attachment of the spectrin–actin cytoskeleton to the erythrocyte plasma membrane. This is accomplished by two specific integral membrane proteins: ankyrin and band 4.1 protein. Ankyrin connects to band 3 protein, which is an anion transport protein in the plasma membrane. Band 4.1 protein binds to glycophorin. Glycophorins (A, B, and C) are transmembrane proteins with several carbohydrate units. Glycophorin A, the major protein marker in the erythrocyte, is a single-pass transmembrane sialoglycoprotein. The anion channels in erythrocytes are important for carbon dioxide transport. (Figure adapted from Luna and Hitt, 1992.)

C. α- and β-Spectrin

Spectrin is the main component of erythrocyte cytoskeletal proteins (**1**). It is a long protein composed of a 260-kDa α chain and a 225-kDa β chain. The chains consist of 20 (α chain) and 18 (β chain) subunits, respectively (**2**), each with 106 amino acids. Each subunit is composed of three α-helical protein strands running counter to one another. The individual subunits are assigned to different domains (I–V in the α chain, and I–IV in the β chain). (Figure adapted from Luna and Hitt, 1992.)

D. Erythrocyte skeletal proteins

Sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis differentiates numerous membrane-associated erythrocyte proteins. Each band of the gel is numbered, and the individual proteins are assigned to them. The main proteins include α- and β-spectrin, ankyrin, an anion-channel protein (band-3 protein), proteins 4.1 and 4.2, actin, and others.

Medical relevance

Inherited red cell membrane disorders are: five types of spherocytosis (182900, 270970) caused by mutations in different genes; elliptocytosis, including the rare pyropoikilocytosis (130500); acanthocytosis (109270); and stomatocytosis type 1 and type 2 (185000). With the exception of an autosomal recessive spherocytosis type 3 (270970), all have autosomal dominant inheritance at loci shown in **D**.

Further Reading

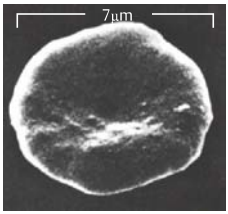
Davies KA, Lux SE. Hereditary disorders of the red cell membrane skeleton. Trends Genet 1989; 5(7): 222–227

Delaunay J. The molecular basis of hereditary red cell membrane disorders. Blood Rev 2007; 21(1): 1–20

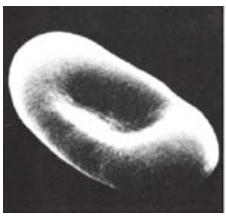
Glader B. Other hereditary Rd blood cell disorders. In: Rimoin DL, et al., eds. Emery and Rimoins's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Luna EJ, Hitt AL. Cytoskeleton--plasma membrane interactions. Science 1992; 258(5084): 955–964

Perrotta S, et al. Hereditary spherocytosis. Lancet 2008; 372(9647): 1411–1426

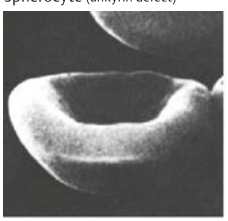


Normal erythrocyte

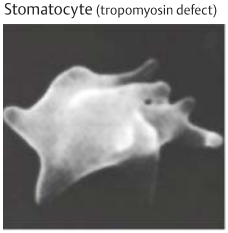


Elliptocyte (spectrin defect)

Spherocyte (ankyrin defect)

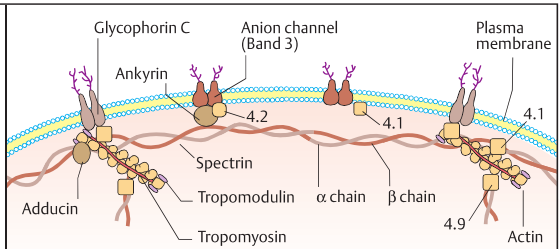


Stomatocyte (tropomyosin defect)

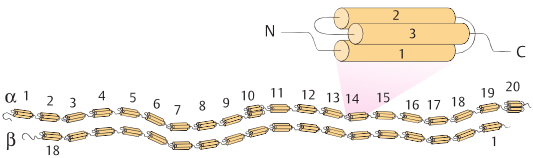


Acanthocyte

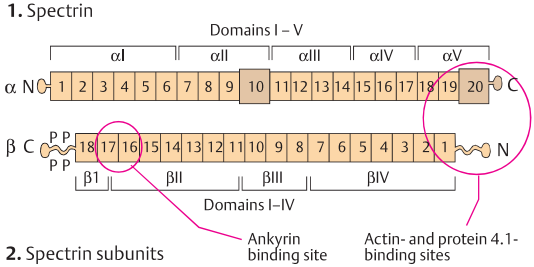
A. Erythrocytes



B. Proteins in the erythrocyte membrane



1. Spectrin



2. Spectrin subunits

C. α- and β-spectrin

Band	SDS gel	Protein	Chromosomal localization	Disease
1		α–Spectrin	1q22 – 25	Elliptocytosis-1
2		β–Spectrin	14q23 – 24	Spherocytosis-2
		Ankyrin	8p11 – 21	Spherocytosis-1
3		Anion channel	17q21 – q22	Acanthocytosis
4.1		Protein 4.1	1q22 – 25	Elliptocytosis-2
4.2		Protein 4.2		
5		Actin	7pter – q22	
6		Glycerol-aldehyde-3-P dehydro-genase	12p13	
7		Tropomyosin (non-muscle)	1q31 – 41	Stomatocytosis

D. Erythrocyte skeleton proteins

Hereditary Muscular Dystrophies

Mutations in genes encoding the many proteins in muscle cause a large group of clinically and genetically different neuromuscular diseases. Hereditary neuromuscular diseases are classified into muscular dystrophies, congenital and other myopathies, spinal muscular atrophies, motor neuron diseases, and others. They are genetically heterogeneous and clinically variable, with well over 50 distinct forms known.

A. The dystrophin–glycan complex

The dystrophin–glycan complex is a system of six interconnected proteins bound to the muscle cell plasma membrane (sarcolemma). The proteins belong to the group of dystroglycans and sarcoglycans. Laminins connect with the extracellular matrix. The central protein is dystrophin, a large elongated protein of 175 nm with a specific structure of two subunits, which are connected to the thin myofilament F-actin (filamentous actin) at the N-terminus and to dystrobrevin and syntrophin at the C-terminus. Dystrophin provides a bridge between the intracellular cytoskeleton involved in the contractile myofilaments and the extracellular matrix. The largest of the interconnecting proteins, α-dystroglycan, is located outside the cell. It is connected to the extracellular matrix by a heterotrimeric protein, laminin-2. Its partner, β-dystroglycan, is embedded in the sarcolemma and connected to a series of other cytoskeletal proteins, which are divided into the sarcoglycan and syntrophin subcomplexes. Two dystrophin molecules connect neighboring dystrophin–glycan complexes.

Several types of congenital muscular dystrophies are known. The complex group of six types of limb girdle muscular dystrophies (OMIM 607155, 253600) is classified according to the type of sarcoglycan involved.

B. Model of the dystrophin molecule

Dystrophin, the largest member of the spectrin superfamily, is composed of 3,685 amino acids, which form four functional domains: (1) the N-terminal actin-binding domain of 336 amino acids; (2) 24 long repeating units, each consisting of 88- to 126-amino-acid triple-helix segments, as in spectrin; (3) a 135-amino-acid cysteine-rich domain, which binds to the sarcolemma proteins; and (4) the C-terminal domain of 320 amino acids with binding sites to syntrophin and dystrobrevin. The triple helix segments form the central rod domain, which is 100 to 125 nm long. (Figure adapted from Koenig et al., 1988.)

C. The dystrophin gene

The human dystrophin gene *(DMD*, 300377) is located on the short arm of the X chromosome in region Xp21.2-p21.1 (**1**). *DMD* is the largest known gene in man, spanning 2.3 million base pairs (2.3 Mb) in 79 exons (**2**). The large DMD transcript has 14 kb. The dystrophin gene contains at least seven intragenic promoters. The primary transcript is alternatively spliced into a variety of different mRNAs that encode smaller proteins expressed in tissues other than muscle cells, especially in the central nervous system. A related gene encodes utrophin (128240).

D. Distribution of deletions

The most frequent types of disease-causing mutations in the *DMD* gene are deletions, which occur in 60 to 65% of patients. They are unevenly distributed. Most frequently involved are exons 43 to 55 and exons 1 to 15, roughly corresponding to the F-actin-binding site and the dystroglycan-binding site. Duplications of one or more exons (in 6% of patients) and point mutations also occur. (Data kindly provided by Professor C.R. Müller-Reible, University of Würzburg, Germany.)

Further Reading

Duchenne muscular dystrophy. Available at: www.ncbi.nlm.nih.gov/books/NBK22263/

Kaplan JC. The 2011 version of the gene table of neuromuscular disorders. Neuromuscul Disord 2010; 20(12): 852–873

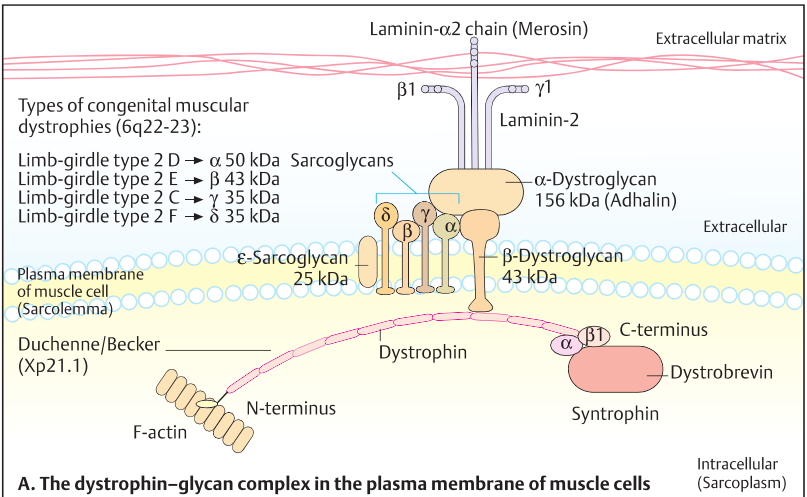
Koenig M, et al. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell 1988; 53(2): 219–228

Mercuri E, et al. Muscular dystrophies. Lancet 2013; 381(9869): 845–860

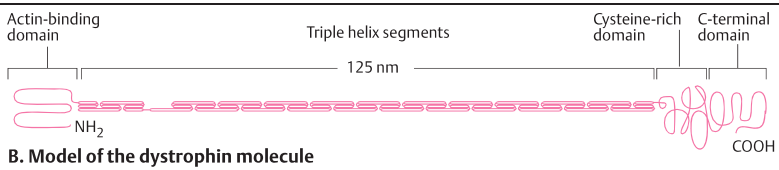
NIH. Muscular dystrophy information page. Available at: www.ninds.nih.gov/Disorders/. Accessed April 7, 2017

OMIM (Online Mendelian Inheritance of Man). Available at: www.ncbi.nlm.nih.gov/omim

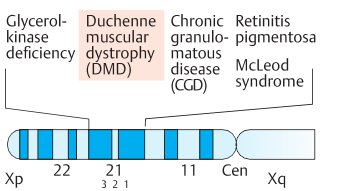
Sarkozy A, et al. Muscular dsytrophies. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013



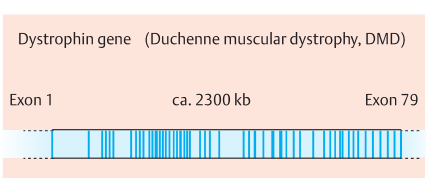
A. The dystrophin–glycan complex in the plasma membrane of muscle cells



B. Model of the dystrophin molecule

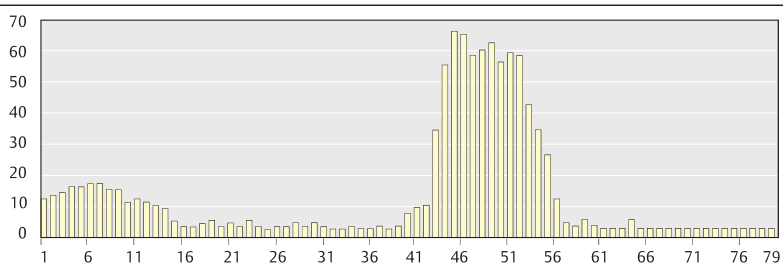


1. Chromosomal location, X chromosome



2. Exon/intron structure and size

C. The dystrophin gene



D. Distribution of deletions in the dystrophin gene

Duchenne Muscular Dystrophy

Duchenne muscular dystrophy (DMD, 310200) is the most common of the muscular dystrophies, with a frequency of 1 in 3,500 liveborn males. It is named after the French neurologist Guillaume Duchenne (1806–1875) who described this disease in 1861. It is caused by mutations in the *DMD* gene (300377), either by a new mutation or by transmission of the mutation from a heterozygous mother. Germline mosaicism has been observed in females carrying a *DMD* mutation in a variable proportion of her germ cells. In approximately one-third of patients, DMA is the result of a new mutation.

A. Clinical signs

Progressive generalized muscle weakness usually results in death from respiratory failure by the age of 18 to 20 years. The age of onset is usually less than 3 to 5 years. Progressive muscular weakness of the hips, thighs, and back causes difficulties in walking and in using steps. A wheelchair is required by the age of 13 to 15 years. Lumbar lordosis and enlarged but weak calves (pseudohypertrophy) are visible (**1**). The affected child performs a characteristic series of maneuvers to rise from a kneeling position (Gower's sign, **2**). Other signs occur during the second decade: cardiomyopathy and impaired cognitive abilities and verbal skills. Female heterozygotes show mild clinical signs in 8%.

A clinically milder variant, Becker muscular dystrophy (BMD, 300376) is an allelic disorder with a milder course and a later age of onset of the disease (mean age, 11). The clinical difference results from different rearrangements of the *DMD* gene. In DMD, the reading frame is changed, whereas in BMD it is maintained. Thus, even a relatively large deletion is compatible with residual muscle function if the reading frame is not changed. (Drawings by Duchenne, 1861, and Gowers, 1879; from Emery and Muntoni, 2003.)

B. Dystrophin analysis in muscle cells

Dystrophin is normally located along the plasma membrane (sarcolemma) of muscle cells (**1**), whereas in patients it is absent (**2**). Female heterozygotes show a patchy distribution of groups of normal and defective muscle cells (**3**) as a result of X inactivation (see p. 184). (Photographs kindly provided by Dr R. Gold, Department of Neurology, University of Würzburg, Germany.)

C. Investigation of a family with DMD

This pedigree is an example of indirect genotyping prior to the direct analysis of the gene available today. A simple two-allele system (marker DXS7) indicates that two patients (III-1 and III-2) carry the allele 1. The mothers, II-1 and II-2, are heterozygous 2–1 for the marker. Their unaffected brother (II-4) carries allele 2. (Figure based on data kindly provided by Dr C. R. Müller-Reible, Würzburg, Germany.)

D. Other forms of muscular dystrophy

The table (see next page) shows examples of other forms of genetically determined muscular dystrophy. Course, diagnosis, and molecular genetic analysis depend on the basic disorder. Selected examples are listed along with their chromosomal location and McKusick (OMIM) number.

Further Reading

Amato AA, et al. Muscular dystrophies and other muscle diseases. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 3487

Bushby A, et al.; DMD Care Considerations Working Group. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. Lancet Neurol 2010; 9(1): 77–93

Bushby K, et al.; DMD Care Considerations Working Group. Diagnosis and management of Duchenne muscular dystrophy, part 2: implementation of multidisciplinary care. Lancet Neurol 2010; 9(2): 177–189

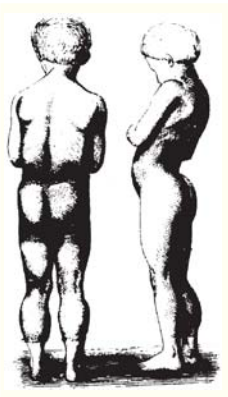
Emery AEH, et al. Duchenne Muscular Dystrophy. 3rd ed. Oxford: Oxford University Press, 2003

Kaplan JC. The 2011 version of the gene table of neuromuscular disorders. Neuromuscul Disord 2010; 20(12): 852–873

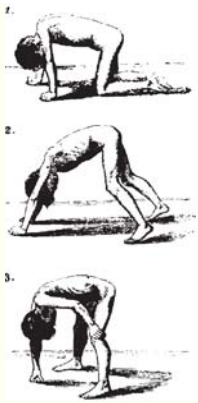
NIH. Muscular dystrophy information page. Available at: www.ninds.nih.gov/Disorders/. Accessed April 7, 2017

Sarkozy A, et al. Muscular dsytrophies. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Worton RG, et al. The X-linked muscular dystrophies. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 5493–5523

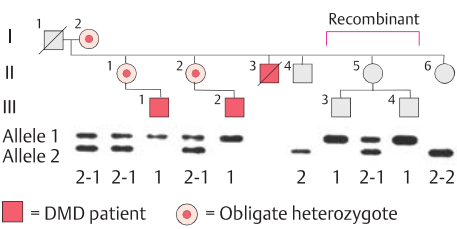


1. Calf hypertrophy and lordosis



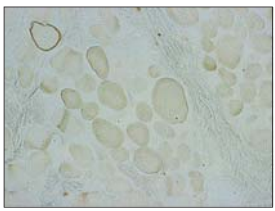
2. Difficulty in rising (Gower's sign)

A. Clinical signs of Duchenne muscular dystrophy



C. Investigation of a family with DMD by DNA markers

1. Normal dystrophin



2. Dystrophin absent

3. Areas lacking dystrophin in heterozygotes

B. Dystrophin analysis in muscle cells

Disease	Chromosomal location	McKusick no.
X-chromosomal:		
Duchenne muscular dystrophy	Xp21.2	310200
Becker muscular dystrophy (allelic with DMD)	Xp21.2	310200
Emery–Dreifuss muscular dystrophy	Xq28	310300
Autosomal dominant:		
Myotonic dystrophy (2 types)	19q13; 3q13	160900; 602668
Facioscapulohumeral dystrophy	4q35 – qter	158900
Oculopharyngeal muscular dystrophy	14q11.2 – q13	164300
Autosomal recessive:		
Duchenne-like muscular dystrophy (LGMD2C)	13q12 – 13	253700
Fukuyama congenital muscular dystrophy	9q31 – 33	253800
Limb-girdle muscular dystrophy (several types)	15q15 – q22, other loci	253600

D. Important forms of hereditary muscular dystrophy in man

FGF Receptor Mutations in Skeletal Dysplasias

Skeletal dysplasias are a large group of clinically and genetically different disorders of bone and cartilage development, often associated with short stature. Mutations in one of three fibroblast growth factor receptors (FGFR1, FGFR2, FGFR) cause a characteristic phenotypic spectrum of skeletal dysplasias. Fibroblast growth factors (FGFs) transmit signals into cells by means of highly specific membrane-bound receptors (FGFRs) that belong to the large family of RTKs (receptor tyrosine kinases, p. 186). Binding of the FGF ligand to their receptors requires both interaction with proteoglycans, such as heparin or heparan sulfate, and dimerization. FGF–FGFR interactions are required to regulate growth in the developing vertebrate limb.

A. Examples of phenotypes

The six examples shown represent distinct types of skeletal dysplasias, which are characterized by their clinical and radiological manifestations. Synostosis of cranial sutures (craniosynostosis) is a typical sign of Crouzon (**1**, 123500), Pfeiffer's (**2**, 101600), and Apert syndromes (**3**, 101200). Complete syndactyly of fingers and toes are a hallmark feature of Apert syndrome. The combination of short stature (mean adult height 115–145 cm), midface hypoplasia, and other skeletal manifestations are characteristic of achondroplasia (**4**, 100800). Thanatophoric dysplasia (**5**, 187600) is a severe, lethal, genetically heterogeneous form. Muenke syndrome (**6**, 602849) is characterized by coronal craniosynostosis (bilateral or unilateral), tarsal and/or carpal fusion, sensorineural hearing loss, and developmental delay. (Photographs 1, 2, and 4–6 courtesy of Dr Maximilian Muenke, National Institutes of Health, United States; photograph 3 from Gilbert, 2016.)

B. Mutations and diseases

Mutations in either of the three fibroblast growth factor receptors (FGFR1, 136350; FGFR2, 134920; FGFR3, 134934) cause the five skeletal dysplasias in different distributions as shown, including Antley Bixler syndrome (OMIM 176943, p. 340, not shown). The extracellular ligand-binding component of FGFRs, three immunoglobulin (Ig)-like domains (IgI, IgII, IgII), are connected to a single transmembrane domain of 22 amino acids. The cytosolic component has two kinase domains. FGFR-related disorders have a rather specific spectrum of allelic gain-of-function dominant mutations. Mutations in either FGFR1 or FGFR2 cause Pfeiffer syndrome. Nearly all individuals (97%) with achondroplasia have the same causative mutation, a G380R substitution (glycine replaced by arginine) within the transmembrane domain. The three clinically distinct disorders achondroplasia, thanatophoric dysplasia, and Muenke syndrome (4–6 in **A**) are allelic since they are all caused by mutations in FGFR3. (Figure modified from Gilbert and Barresi, 2016.)

C. Effects on the growth plate

The normal pattern of chondrogenesis, i.e., the normal transition of resting chondrocytes to proliferating chondrocytes, and early osteoblasts in the growth plate (left) of developing bone is totally disorganized thanatophoric dysplasia with lack of proliferating chondrocytes (right). Most FGFR mutations increase the affinity of the ligand to the receptor (gain-of-function). For further information see Table 15, Appendix p. 414.

Further Reading

Bonafe L, et al. Nosology and classification of genetic skeletal disorders: 2015 revision. Am J Med Genet A 2015; 167A(12): 2869–2892

Chen C, et al. Skeleton Genetics: a comprehensive database for genes and mutations related to genetic skeletal disorders. Database (Oxford) 2016, 2016: baw127

Gilbert SF, Barresi MJF. Developmental Biology. 11th ed. Sunderland: Sinauer, 2016

Givol D, et al. Molecular and cellular biology of growth factor signaling pathway. In: Erickson RP, et al., eds. Epstein's Inborn Errors of Development. 3rd ed. Oxford: Oxford University Press, 2016

Muenke M, ed. Craniosynostoses. Molecular Genetics, Principles of Diagnosis, and Treatment. Basel: Karger, 2011

Offiah AC. Skeletal dysplasias: An overview. Endocr Dev 2015; 28: 259–276

Wilkie ACM. FGF receptor mutations: bone dysplasia, craniosynostosis, and other syndromes. In: Erickson RP, et al., eds. Epstein's Inborn Errors of Development. 3rd ed. Oxford: Oxford University Press, 2016

Mother and child
1. Crouzon syndrome

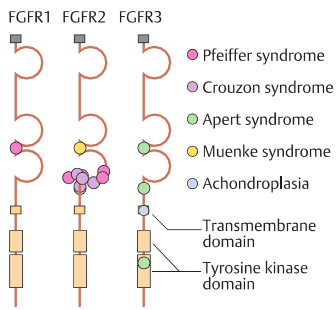
Mother and child affected
2. Pfeiffer syndrome

3. Apert syndrome
A. Examples of phenotypes

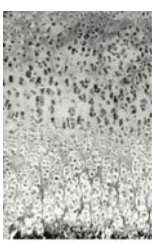
4. Achondroplasia

5. Thanatophoric dysplasia

6. Muenke syndrome



B. Mutations and diseases



Normal

Thanatophoric dysplasia

C. Effects on the growth plate

Marfan and Loeys–Dietz Syndromes

Marfan syndrome (MFS) and Loeys–Dietz syndrome (LDS) are autosomal dominant syndromes associated with aortic aneurysms and other manifestations. MFS, first described in 1896, is caused by mutations in the fibrillin-1 gene *FBN1* (134797). Mutations in a similar gene, *FBN2* (612570), located at 5q23-q31, cause a related disorder known as congenital contractural arachnodactyly (CCA, 121050).
LDS, first described in 2005, is caused by mutations either in the *TGFBR1* gene (transforming growth factor beta receptor 1, 190181) in LDS type 1, or in the *TGFBR2* gene (190182) in LDS type 2.

A. Marfan syndrome (MFS)

MFS (154700) mainly affects the skeleton, the cardiovascular system, and the eye. Hyperextensible joints (**1**), long fingers (arachnodactyly, **2**), pectus excavatum (**3**), subluxation of the lens in 50 to 80% of cases (**4**) are leading signs. The main cardiac manifestation is progressive aortic root dilatation and a high risk for aortic aneurysms leading to aortic dissection. β-Adrenergic blockade (propranolol) delays the rate of aortic dilatation. In view of excessive signaling by TGFβ cytokines, blocking the angiotensin II type 1 receptor (AGTR1; OMIM 106165) promises to be an effective therapeutic approach. (Photographs in part provided by Dr. Beate Albrecht, Essen, and Dr. M. Siepe, Herzzentrum Freiburg, Germany.)

B. Loeys–Dietz syndrome (LDS)

The main manifestations of LDS (609192) are facial dysmorphia with hypertelorism, pectus carinatum, scoliosis (**1**), deformed feet (**2**), vascular tortuosity (**3**), cleft palate or bifid uvula in some patients (**4**) associated with mitral valve prolapse, aneurysms in the aorta (**5**) and other arteries. The phenotype may also overlap with Shprintzen–Goldberg syndrome (182212). (Photographs from Lindsay and Dietz, 2011.)

C. Fibrillin-1 protein (FBN1)

Fibrillins are glycoproteins in the extracellular microfibrils in connective tissue. The FBN1 protein consists of several distinct regions. There are two types of motif: 43 calcium-binding EGF-like (epidermal growth factorlike) motifs, and seven motifs, each with eight cysteine residues; there are two hybrid domains consisting of EGF-like motifs and motifs containing eight cysteine residues. The protein is encoded by the *FBN1* gene located at 15q21.1, consisting of 65 exons extending over 235-kb genomic DNA. The mRNA is 10 kb in length (9,749 nucleotides) with three alternatively spliced noncoding 5′ exons.

D. *TGFBR1* and *TGFBR2* genes

These genes encode a serine/threonine kinase transmembrane receptor for TGFβ. Activation by their ligand regulates cell proliferation and differentiation in various developmental processes. Mutations in these two genes lead to increased TGFβ signaling and cause Loeys–Dietz syndrome. (Figure adapted from Loeys et al., 2006.)

E. Increased TGF-β signaling

TGFβ signaling stimulates multiple intracellular pathways. One of these involves extracellular matrix proteins, including fibrillin-1. Inhibition of increased TGFβ signaling ameliorates aortic aneurysms in mice. (Figure based on Habashi et al., 2011.)

Further Reading

Aubart M, et al. The clinical presentation of Marfan syndrome is modulated by expression of wild-type FBN1 allele. Hum Mol Genet 2015; 24(10): 2764–2770

Habashi JP, et al. Angiotensin II type 2 receptor signaling attenuates aortic aneurysm in mice through ERK antagonism. Science 2011; 332(6027): 361–365

Judge DP, Dietz HC. Marfan's syndrome. Lancet 2005; 366(9501): 1965–1976

Lindsay ME, Dietz HC. Lessons on the pathogenesis of aneurysm from heritable conditions. Nature 2011; 473(7347): 308–316

Loeys BL, et al. Aneurysm syndromes caused by mutations in the TGF-β receptor. N Engl J Med 2006; 355(8): 788–798

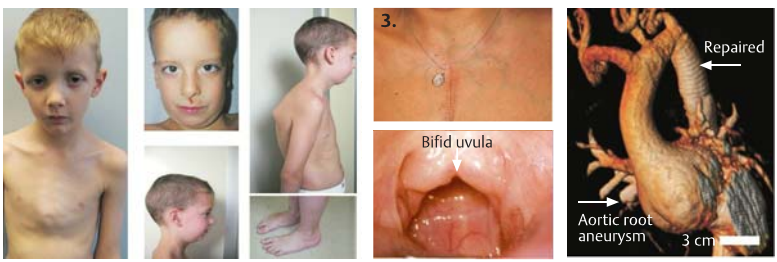
MacCarrick G, et al. Loeys-Dietz syndrome: a primer for diagnosis and management. Genet Med 2014; 16(8): 576–587

Pyeritz RE. Marfan syndrome and related disorders. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Stheneur C, et al. Study of phenotype evolution during childhood in Marfan syndrome to improve clinical recognition. Genet Med 2014; 16(3): 246–250

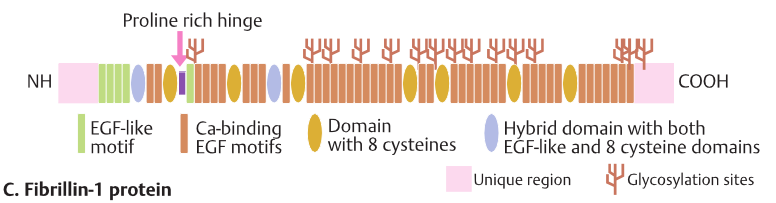
1. 2. 3. 4.

A. Marfan syndrome

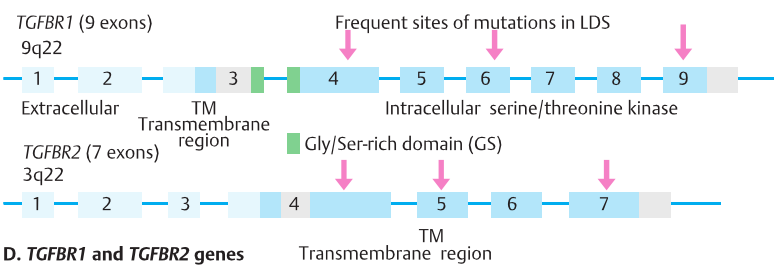


1. 2. 4. 5.

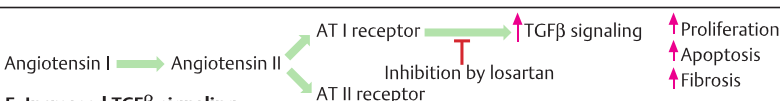
B. Loeys-Dietz syndrome



C. Fibrillin-1 protein



D. *TGFBR1* and *TGFBR2* genes



E. Increased TGFβ signaling

Collagen Molecule Disorders

Numerous genetic disorders result from mutations in genes encoding different forms of collagen molecules. They consist of a large group of insoluble, extracellular glycoproteins with specialized functions in the maintenance of tissue shape and structure. Collagens account for approximately one-quarter of the total body protein, occurring in skin, bones, tendons, cartilage, ligaments, blood vessels, teeth, basement membranes, and supporting tissues of the internal organs. Most collagens have a remarkable mechanical resilience. The 27 known types of collagens, numbered I to XXVII, are classified according to their main structural and functional features. The most important collagens in humans are fibrillar collagen types I, II, III, V, and XI, and basal membrane collagen (type IV). Collagen types I, II, and III account for 80 to 90% of collagens. Type I is the main component of tendons, ligaments, and bones; type II occurs in cartilage and the notochord of vertebrate embryos; type III occurs in arteries, the intestine, and the uterus; and type IV, in the basal lamina of epithelia, in particular around the glomeruli of the kidney, and in blood capillaries.

A. Collagen structure

Collagen is a triple helix of three chains. It is synthesized in seven principal steps. The sequence of the approximately 1000 amino acids of a typical collagen is simple and periodic (**1**). Every third amino acid is glycine (Gly). Other amino acids alternate between the glycines. The general structural motif is $(Gly–X–Y)_n$. X is either proline or hydroxyproline; Y is either lysine or hydroxylysine (**2**). Three chains of procollagen form a triple helix (**3**). In collagen type I, the helix is composed of two identical α1 chains and one α2 chain. First a precursor molecule, *procollagen* is formed (**4**). Procollagen peptidases cleave peptides at the N-terminal and C-terminal ends to form *tropocollagen* (**5**). Tropocollagen molecules, connected by numerous hydroxylated proline and lysine residues, form a *collagen fibril* (**6**). Each fibril consists of staggered, parallel rows of end-to-end tropocollagen molecules, separated by gaps (**7**). Collagen fibrils are visible as transverse stripes under the electron microscope (EM) (**8**). A fiber of 1-mm diameter can hold a weight of almost 10 kg. (Photograph from Stryer, 1995.)

B. Prototype of a gene for procollagen

Procollagen type II consists of a triple helix of three procollagen type II chains, designated as α1[II]. The corresponding gene, *COL2A1* (120140), located at 12q13.11, consists of 52 exons of different sizes, each encoding 5, 6, 11, 12, or 18 Gly–X–Y units. The translated part of exon 1 (85 bp) encodes a signal peptide necessary for secretion. The genes for procollagen types I, II, and III differ in that some exons are fused, but otherwise they are similar, especially for the three main fibrillar collagen types (I, II, III). Mutations in *COL2A1* occur in several different disorders (see OMIM 120140).

C. Gene structure of procollagen type α1(I)

Procollagen type I consists of two α1 chains and one α2 chain, designated $\alpha1(I)_2\ \alpha2(II)$. The corresponding genes are *COL1A1* (120150) and *COL1A2* (120160). The 52 exons of the gene *COL1A1* encoding procollagen type I correspond to the different domains (A–G) of procollagen α1 (I). The *COL1A2* gene is approximately twice as large (approximately 40 kb) as the *COL1A1* gene because the introns between the exons are twice as long as in *COL1A1*.

Examples of human diseases caused by mutations in one of the genes encoding different collagen molecules are listed in the Appendix, Table 16, p. 415.

Further Reading

Bateman JF, et al. Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. Nat Rev Genet 2009; 10(3): 173–183

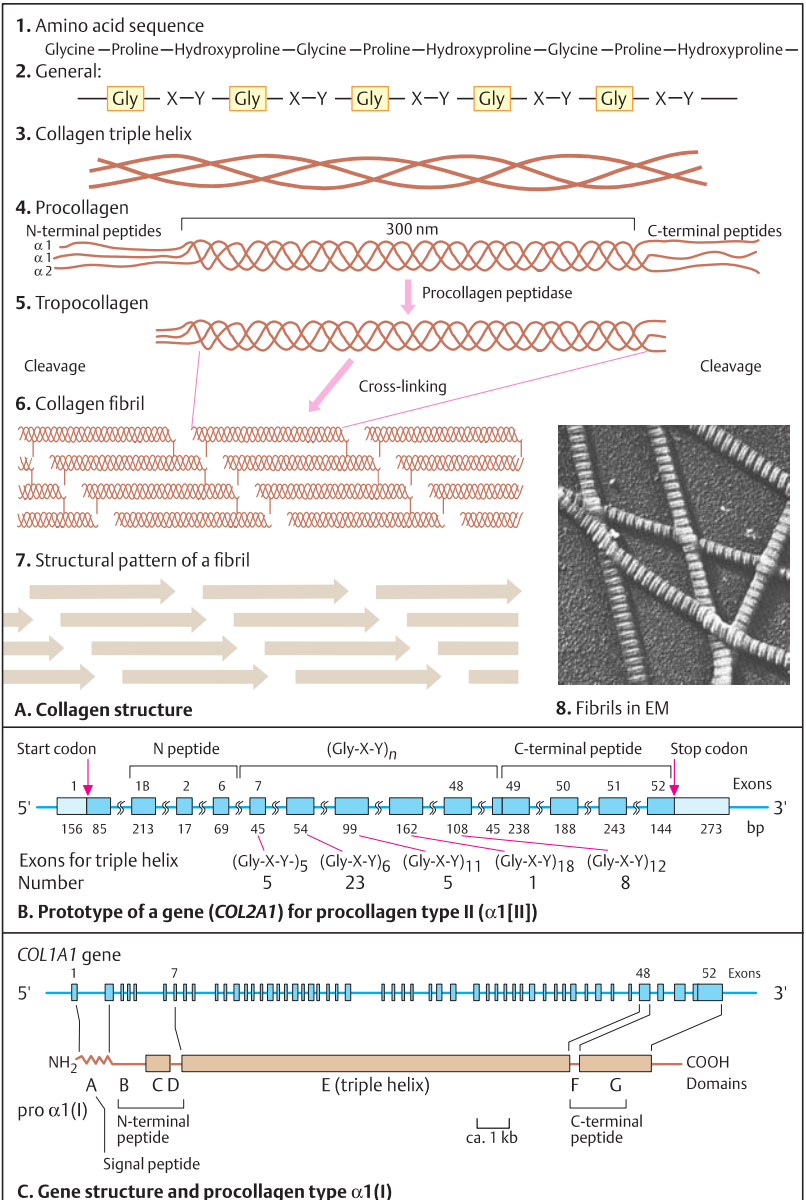
Byers PH. Disorders of collagen synthesis and structure. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 5241–5285

Chu M-L, Prockrop DJ. Collagen gene structure. In: Royce PM, et al., eds. Connective Tissue and Its Heritable Disorders. 2nd ed. New York, NY: Wiley-Liss, 2002: 149–165

Munns C, Sillence D. Osteogenesis imperfecta (and other disorders of bone matrix. Chapter 156. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Rauch F, et al. Osteogenesis imperfecta. Lancet 2004; 363(9418): 1377–1385

Stryer L. Biochemistry. 4th ed. New York, NY: W. H. Freeman, 1995


1. Amino acid sequence
Glycine—Proline—Hydroxyproline—Glycine—Proline—Hydroxyproline—Glycine—Proline—Hydroxyproline—
2. General:
Gly X—Y Gly X—Y Gly X—Y Gly X—Y Gly X—Y
3. Collagen triple helix
4. Procollagen
N-terminal peptides
300 nm
C-terminal peptides
α 1
α 1
α 2
Procollagen peptidase
5. Tropocollagen
Cleavage
Cleavage
Cross-linking
6. Collagen fibril
7. Structural pattern of a fibril
A. Collagen structure
8. Fibrils in EM
Start codon
N peptide
(Gly-X-Y)n
C-terminal peptide
Stop codon
Exons
5'
3'
1 1B 2 6 7 48 49 50 51 52
156 85 213 17 69 45 54 99 162 108 45 238 188 243 144 273 bp
Exons for triple helix
Number
(Gly-X-Y-)5 (Gly-X-Y)6 (Gly-X-Y)11 (Gly-X-Y)18 (Gly-X-Y)12
5 23 5 1 8
B. Prototype of a gene (COL2A1) for procollagen type II (α1[II])
COL1A1 gene
1 7 48 52 Exons
5' 3'
NH2
COOH
Domains
A B C D E (triple helix) F G
pro α1(I)
N-terminal peptide
Signal peptide
ca. 1 kb
C-terminal peptide
C. Gene structure and procollagen type α1(I)

Osteogenesis Imperfecta

Osteogenesis imperfecta (OI, 120150), or "brittle bone disease," is a heterogeneous group of approximately 20 clinically and genetically different types of diseases with a total frequency of approximately 1 in 10000 individuals (phenotypic series in OMIM 166200). Common manifestations include spontaneously occurring bone fractures, bone deformity, small stature, defective dentition (dentinogenesis imperfecta), hearing impairment due to faulty formation of the auditory ossicles, and blue sclerae. In most, but not all patients, collagen type I is defective owing to a mutation in one of the two genes encoding the two chains *COL1A1* (120150) or *COL1A2* (120160). Both autosomal dominant and recessive types of OI exist.

A. Mutations in collagen

In normal procollagen type I the two chains pro α1(I) and α2(I) are produced in equal amounts (**1**). Some mutations lead to reduced production of pro α1(I) (**2**). The resulting imbalance causes degradation of α2(I) chains. Thus, the amount of procollagen is reduced. Different types of mutations in the genes for pro α1 (I) and α2(I) *(COL1A1* and *COL1A2)* cause defective procollagen (**3**) because of a deletion in a *COL1A1* allele, either due to a splicing defect or other type of mutation. Mutations in the *COL1A1* gene are more severe than mutations in the COL1*A2* gene because more defective collagen is formed with the former. (Figure adapted from Wenstrup et al., 1990.)

B. Mutations and phenotype

A relationship between genotype and phenotype exists. Generally, mutations in the 3′ region are more serious than mutations in the 5 region (*position effect*). Mutations of the pro α1 (I) chain are more severe than those in the pro α2(I) chain (*chain effect*). The substitution of a larger amino acid for glycine, which is indispensable for the formation of the triple helix, leads to a severe disorder (*size effect*). Different types of mutations may occur, such as deletions, mutations in the promoter or enhancer, and splicing mutations. The codons (AAG, AAA) for the amino acid lysine, which occurs frequently in collagen, are readily transformed into a stop codon by substitution of the first adenine by a thymine (TAG or TAA), so that a short, unstable procollagen is formed. Splicing mutations may lead to the loss of exons (exon skipping). (Figure adapted from Byers, 2001.)

C. Different types of OI

OI is classified into four basic phenotypes, I to IV, according to the Sillence classification. OI types I and IV (**1**, 166220), progressively deforming type III (**2**), and neonatal lethal OI type II (**3**) with distinctly thickened and shortened long bones in the newborn). Germline mosaicism has been shown to account for rare instances of affected siblings being born to unaffected parents.

Several new types of OI have been described: types V to XVII (see OMIM 166200), some with hyperplastic callus formation. About 8 types are autosomal dominant and 10 autosomal recessive (OMIM 166220).

Further Reading

Basel D, Steiner RD. Osteogenesis imperfecta: recent findings shed new light on this once well-understood condition. Genet Med 2009; 11(6): 375–385

Byers PH et al. Genetic evaluation of suspected osteogeneisi imperfecta (OI). Genet Med 2006; 8: 383–388

Byers PH, et al. Osteogenesis imperfecta. In: Royce PM, et al., eds. Connective Tissue and Its Heritable Disorders. 2nd ed. New York, NY: Wiley-Liss, 2002: 385–430

Byers PH. Disorders of collagen synthesis and structure. In: Scriver CR, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 5241–5285

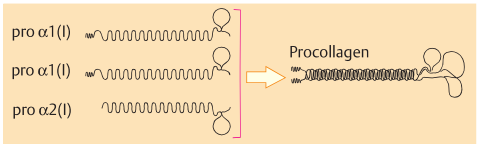
Forlino A, et al. New perspectives on osteogenesis imperfecta. Nat Rev Endocrinol 2011; 7(9): 540–557

Munns C, et al. Osteogenesis imperfecta (and other disorders of bone matrix. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

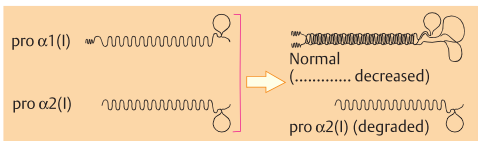
Sillence DO, et al. Genetic heterogeneity in osteogenesis imperfecta. J Med Genet 1979; 16(2): 101–116

Wenstrup RJ, et al. Distinct biochemical phenotypes predict clinical severity in nonlethal variants of osteogenesis imperfecta. Am J Hum Genet 1990; 46(5): 975–982

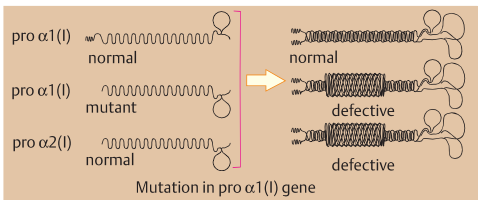
Willaert A, et al. Recessive osteogenesis imperfecta caused by LEPRE1 mutations: clinical documentation and identification of the splice form responsible for prolyl 3-hydroxylation. J Med Genet 2009; 46(4): 233–241

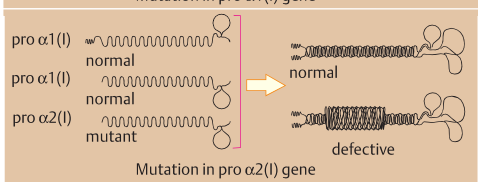


1. Normal



2. Decreased synthesis of procollagen α1(I)

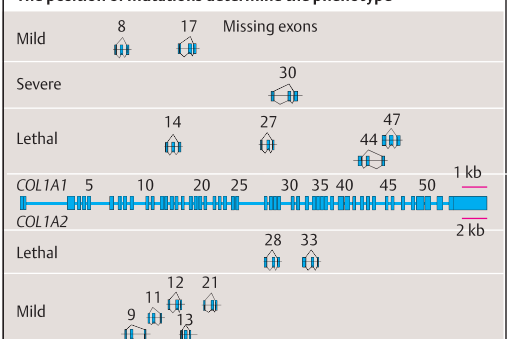




3. Defective procollagen due to a mutation

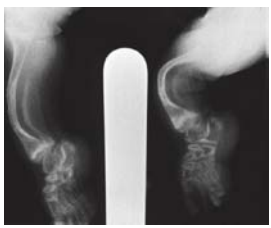
A. Molecular mechanisms in osteogenesis imperfecta

The position of mutations determine the phenotype



B. Mutations and phenotype

1. Bone deformation (OI type I)



2. Severe deformation (OI type III)

3. Fatal form (OI type II)

C. Different forms of osteogenesis imperfecta

Molecular Basis of Bone Development

Bone develops from three mesodermal cell lineages committed to differentiate into three specialized cell types: *chondrocytes* (cartilage-forming cells), *osteoblasts* (bone-forming cells), and *osteoclasts* (bone-degrading cells). Two major processes form bone (osteogenesis): (1) direct conversion of mesenchymal tissue into bone tissue (*intramembranous or dermal ossification*); and (2) *endochondral ossification*, with cartilage intermediates produced by chondrocytes, which are later replaced by bone cells (osteoblasts). Osteoblasts produce most of the proteins for the extracellular bone matrix and control its mineralization. The osteoblast cell lineage involves osteoblast-specific transcription factors (OSFs). One such transcription factor is a major regulator of osteoblast differentiation in direct intramembranous bone formation: the core-binding factor CBFA1 (runt-related transcription factor, RUNX2; OMIM 600211). Three mammalian *Runx* genes are homologs of the Drosophila pair-rule gene *Runt*.

A. Effects of homozygous *Runx2* mutations in mice

Targeted disruption of the *Runx2* gene, located on mouse chromosome 17, in the homozygous state (**–/–**) results in lack of ossification of the entire skeleton (**1**). Normal calcified skeleton stains red with alizarin (a, c), whereas bones lacking ossification stain blue (b, d, e). These mice are small and die at birth from respiratory failure. The humerus and humeral tuberosity (circle, c) in heterozygous mice (+/**–**, d) and homozygous mice (**–/–**) show reduced ossification of the long bones and severe hypoplasia, respectively (d, e). The skull and thorax are also severely affected (**2**). Heterozygous (+/**–**) mice lack ossification of the skull (b). Normal calcified bone is stained red by alizarin red, here at embryonic day 17.5, three and a half days before birth. Cartilage is stained blue by alcian blue. Heterozygous mice lack clavicles (arrows, d) in contrast to normal mice (c).

B. Cleidocranial dysplasia in humans

Cleidocranial dysplasia (CCD, 119600) is an autosomal dominant skeletal disease caused by mutation in the human *RUNX2* gene (600211), localized at 6p21.1. It is characterized by absence of the clavicles and deficient bone formation of the skull. Radiological findings show generalized underossification. Patients can oppose their shoulders (**1**) due to absence of the clavicles (**2**; photograph by Dr J. Warkany, Children's Hospital Research Foundation, Cincinnati, United States). The calvarium (skull case) is enlarged, with a poorly ossified midfrontal area (**3**).

C. The human *RUNX2* gene

The *RUNX2* gene at 6p21 encodes a transcription factor of the core-binding factor (CBF) Runt-related family (OMIM 600211). It has nine exons (not seven as previously determined and shown). It contains two alternative transcription initiation sites with two promoters, P1 and P2. Part of exons 1, 2, and 3 encode the DNA-binding *runt* domain; exons 4, 5, 6, and 7 encode the transcriptional activation and repression domains. The nuclear localization signal (NLS) is located at the 5' end of exon 3. Exon 6 is alternatively spliced and unique to *RUNX2*. The role of the *RUNX2* gene also includes a major regulatory function in chondrocyte differentiation during endochondral bone formation. As such, it functions as a "master gene" in bone development. All mutations result in loss of function, that is, haploinsufficiency causes the CCD phenotype. (Figures in A and B kindly provided by Professor Stefan Mundlos, Berlin, Germany.)

Further Reading

Dinçsoy Bir F, et al. Cleidocranial dysplasia: Clinical, endocrinologic and molecular findings in 15 patients from 11 families. Eur J Med Genet 2017; 60 (3): 163–168

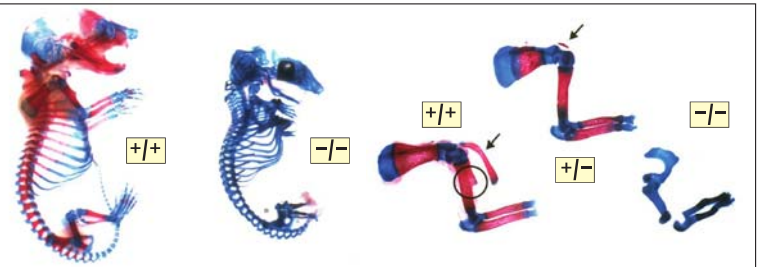
Mundlos S. Cleidocranial dysplasia: clinical and molecular genetics. J Med Genet 1999; 36(3): 177–182

Mundlos S, et al. Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. Cell 1997; 89(5): 773–779

Superti-Furga A, et al. Nosology and classification of genetic skeletal disorders: 2006 revision. Am J Med Genet A 2007; 143A(1): 1–18

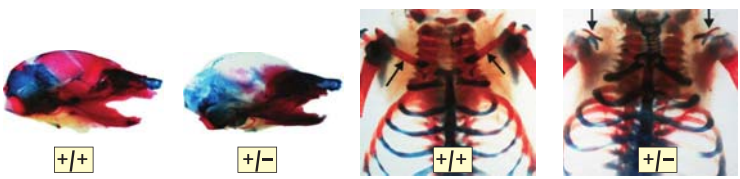
Zaidi SK, et al. Runx2 deficiency and defective subnuclear targeting bypass senescence to promote immortalization and tumorigenic potential. Proc Natl Acad Sci U S A 2007; 104(50): 19861–19866

Zheng Q, et al. Dysregulation of chondrogenesis in human cleidocranial dysplasia. Am J Hum Genet 2005; 77(2): 305–312



a) Normal skeleton **b)** Homozygous mutant, lack of ossified bone **c)** Normal humerus **d)** Hypoplasia **e)** Lack of bone development

1. Skeleton and humerus



a) Normal skull **b)** Reduced mineralization **c)** Normal thorax **d)** Absent clavicles, hypoplastic humoral tuberosity

2. Skull and thorax

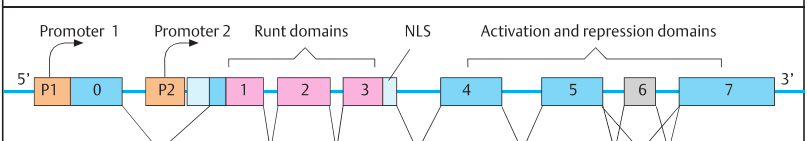
A. Effect of *Runx2* mutations in mice

1. Absent clavicles

2. X-ray: narrow thorax and absent clavicles

3. Lack of skull ossification

B. Cleidocranial dysplasia in humans



C. The human *RUNX2* gene

Hemoglobin

Hemoglobin is a tetrameric oxgen-transporting molecule, named by Hoppe-Seyler in 1862. It consists of two slightly different pairs of polypeptide globin chains with a characteristic three-dimensional structure. Each globin chain contains an oxygen-binding site. Hemoglobin was the first human protein to be defined by its amino acid sequence by Ingram in 1956 and its three-dimensional structure at 5.5-Å resolution by Perutz et al., in 1960. The genes encoding the individual globin chains were the first genes for which the molecular structure was elucidated in 1979. More than 1,000 genetic variants are known, and many cause severe hemoglobin disorders.

A. Types of hemoglobin

Six different globin polypeptide chains form different hemoglobin tetramers at different developmental stages. The subunits of adult hemoglobin are designated by the Greek letters α, β, γ, and δ. Specialized globin subunits exist in mammals during embryonic development, designated as ε (epsilon) and ζ (zeta). A hemoglobin molecule is stable only when the two pairs of globin differ. Adult hemoglobin (HbA) is composed of two α and two β chains (α2β2), and fetal hemoglobin (HbF) is composed of two α and two γ chains (α2γ2). A minor adult hemoglobin, HbA_2, has a variant delta chain (α2δ2). Different types of globin chains are synthesized during embryonic, fetal, and postnatal life. (Figure adapted from Lehmann and Huntsman, 1974.)

B. Hemoglobins in thalassemias

The thalassemias are a group of genetically determined disorders of hemoglobin synthesis. They are classified according to which chain is affected, the α chain or the β chain. Hemoglobins with four identical globin chains are highly unstable and incompatible with life (HbH with four β chains [β4], Hb Bart's with four γ chains [γ4]). (Figure adapted from Lehmann and Huntsman, 1974.)

C. Evolution of hemoglobin

The different types of hemoglobin existing in mammals today arose by duplication of genes encoding the globin chains. A primordial single-chain oxygen-binding protein, present early in evolution, adapted to the rising concentration of oxygen in the earth's atmosphere at an early stage approximately 1100 million years ago. Subsequent duplications gave rise to the several different types of globin genes. Approximately 500 million years ago, a hemoglobin gene underwent the first duplication, and from this the genes encoding the α type and the β type hemoglobins arose. During the past 100 million years, hemoglobin evolved from an ancestral β type into different types: δ, two types of γ (γA and γG), and ε, by further duplication events.

D. Globin synthesis during embryonic development

During early embryonic stages transient embryonic hemoglobins are present: Hb Gower 1 (ζ2ε2), Hb Gower 2 (α2ε2), and Hb Portland (ζ2γ2). (Most hemoglobins are named after the town or institution where they were discovered.) During the first 6 weeks in humans the embryonic hemoglobins in the yolk sac, from the 6th week until prior to birth (fetal hemoglobin HbF, α2γ2), in the liver and spleen, and during adult life in red blood cell precursors in the bone marrow. They differ in oxygen-binding affinity. Thus, oxygen delivery is optimized for different phases of development. (Figure adapted from Weatherall et al., 2001.)

Further Reading

Forget BG, et al. Classification of the disorders of hemoglobin. Cold Spring Harb Perspect Med 2013; 3 (2): a011684

Ingram VM. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. Nature 1956; 178(4537): 792–794

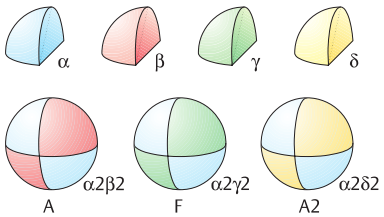
Lehmann H, Huntsman RG. Man's Hemoglobins. Amsterdam: North-Holland, 1974

Old J. Hemoglobinopathies and Thalassemias. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

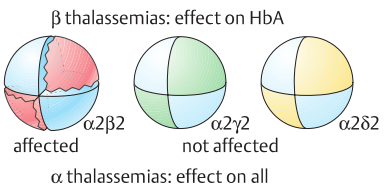
Perutz MF, et al. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature 1960; 185 (4711): 416–422

Smith EC, Orkin SH. Hemoglobin genetics: recent contributions of GWAS and gene editing. Hum Mol Genet 2016; 25(R2): R99–R105

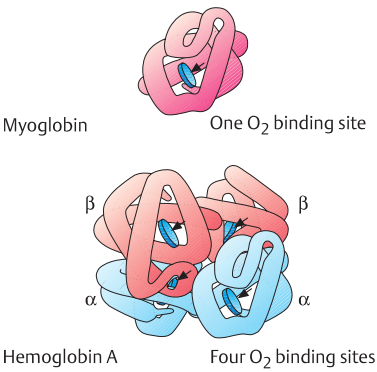
Weatherall DJ, et al. The hemoglobinopathies. In: Scriver CR, et al. eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 179–190

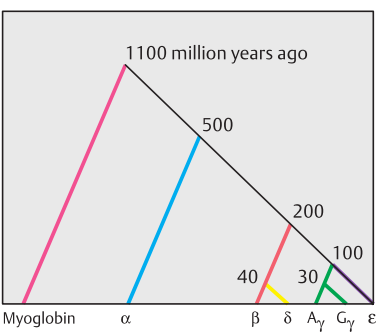


A. Types of hemoglobin

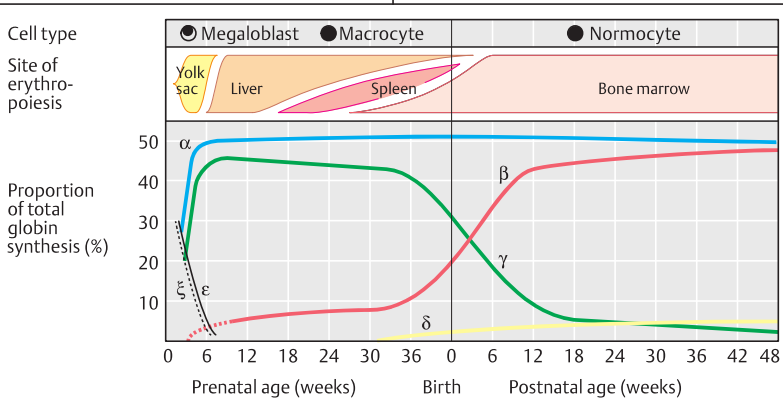


B. Hemoglobin in thalassemias





C. Evolution of hemoglobin



D. Globin synthesis during ontogeny

Hemoglobin Genes

The different hemoglobin chains are encoded by clusters of α-type (two α and one ζ and β-type globin genes (β, γG, γA, δ, and ε) on human chromosomes 16 and 11, respectively. A specific gene is responsible for each type of different globin polypeptide chain. They are arranged and expressed in a sequence, according to the time of activation during developmental stages. Their structural and DNA sequence similarity predisposes to malalignment during meiosis and unequal crossing-over leading to duplications and deficiencies.

A. The β- and α-globin genes

The β-globin-like genes (β, δ, γA, γG, and ε) of man are located at the Hb β locus HBB (OMIM 141900) as a gene cluster in the 3′ to 5′ direction on the short arm of chromosome 11 in region 1, band 5.5 (11p15.5). They span approximately 55,000 bp, or 55 kb (kilobases), of DNA (**1**). There are two γ genes, γG and γA, which have arisen by a duplication event. They differ only in codon 136, which encodes an alanine in γA and a glycine in γG. A pseudogene (ψβ1) containing deletions and internal stop codons is located between the γA gene and the δ gene. A locus control region (LCR) located upstream (in the 5′ direction) jointly regulates these genes.

Two α-globin genes are located at the α locus HBA1 (OMIM 141800) and HBA2 (OMIM 141850) on the short arm of human chromosome 16 (16pter-p13.11) on a DNA segment of approximately 35 kb. A ζ gene, which is active during the embryonic period only, lies in the 5′ direction. Three pseudogenes (ψζ2, ψζ1, ψα2, and ψα1) are located 5′ of the α genes. A further gene locus *HBQ1* (theta, Θ; 142240), with possible function in early erythroid tissue function, has been identified in this region.

The β-globin gene, the prototype, spans approximately 1.6 kb (1,600 bp). It consists of three exons separated by a short and a long intron (2). The coding sequences of the other β-like genes are also arranged in three exons. The α genes *HBA1* and *HBA2* span approximately 0.8 kb (800 bp).

B. Tertiary structure of the β-globin chain

The characteristic three-dimensional structures of myoglobin and of the hemoglobin α and β chains, shown schematically, are very similar, although their amino acid sequences correspond in only 24 of 141 positions. The β chain, with 146 amino acids, is somewhat longer than the α chain, with 141 amino acids. Each chain harbors a heme group with an oxygen-binding site inside the molecule, protected from the surroundings. (Figure adapted from Weatherall et al., 2001.)

C. Functional domains of the β chain

Three functional and structural domains can be distinguished in all globin chains. They correspond to the three exons of the gene. Two domains, consisting of amino acids 1 to 30 and 105 to 146 (encoded by exons 1 and 3), are located on the outside of the molecule, and consist of mainly hydrophilic amino acids. A third domain, lying inside the molecule (encoded by exon 2), contains the oxygen-binding site and consists of mainly nonpolar hydrophobic amino acids. The hydrophilic amino acids of the two chains forming the outside of the molecule render it flexible and water soluble. (Figure adapted from Gilbert, 1978.)

Further Reading

Antonarakis SE, et al. DNA polymorphism and molecular pathology of the human globin gene clusters. Hum Genet 1985; 69(1): 1–14

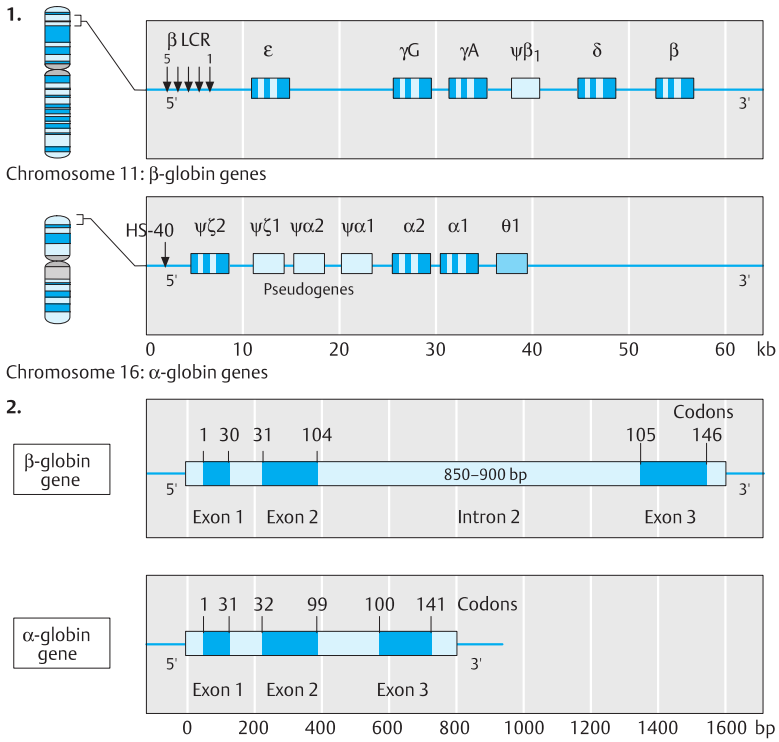
Gilbert W. Why genes in pieces? Nature 1978; 271 (5645): 501–502

Stamatoyannopoulos G, et al. The Molecular Basis of Blood Disease. 4th ed. New York, NY: Saunders, 2001

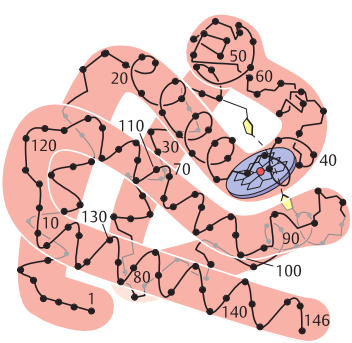
Weatherall DJ, et al. The hemoglobinopathies. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. New York, NY: McGraw-Hill, 2001: 179–190. Available at: https://ommbid.mhmedical.com/. Accessed May 4, 2017

Weatherall DJ, et al. The Thalassaemia Syndromes. 4th ed. Oxford: Blackwell Science, 2001

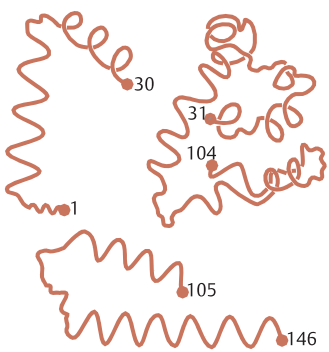
Weatherall DJ. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. Nat Rev Genet 2001; 2(4): 245–255



A. The β- and α-globin genes



B. Tertiary structure of the β-globin chain



C. Three domains of the β chain

Sickle Cell Disease

Sickle cell disease, first described in 1910 by Herrick, is a severe hemolytic anemia associated with many complications. It is caused by a structurally abnormal hemoglobin (HbS) resulting from homozygosity for a specific mutation in the β-globin gene. It is frequent in tropical regions where malaria is endemic (see p. 126). With a frequency of 1 in 500, it is an important cause of morbidity and mortality in these regions. Sickle cell anemia was the first human disease to be understood at the molecular level by Pauling et al. in 1949. It is transmitted by autosomal recessive inheritance (Neel, 1949). Heterozygous carriers can readily be identified.

A. Sickle cells

The leading hematological manifestation is anemia associated with mainly sickle-shaped erythrocytes in a blood smear under the light microscope of affected persons (**1**). Sickle cells lose the pliability needed to pass through small capillary blood vessels and provoke occlusions of small blood vessels (**2**). As a result, the blood supply to many tissues is diminished. In a normal blood smear (**3**), erythrocytes appear as regular round disks of approximately 7 µm in diameter. In the course of the disease, acute hemolytic episodes, called sickle crises, occur. Heterozygotes (sickle trait) show occasional sickle cells but do not develop sickle crises, and are associated with only very mild signs and symptoms. Other sickle hemoglobinopathies occur when the HbS mutation is present on one allele and a different mutation at the same site is present on the other allele. (Figure in 1 and 3, Dr Daniel Nigro, Santa Ana College, California, United States; 2, National Heart, Lung and Blood Institute, Bethesda, Maryland, United States.)

B. Consequences of the sickle cell mutation

All manifestations of sickle cell anemia can be understood on the basis of the underlying mutation. In 1956, V.M. Ingram determined, by using amino acid sequence analysis, that glutamic acid is replaced by valine in codon 6 of the β-globin gene (mutation E6V). This alteration results from the transversion of an adenine (A) to a thymine (T), which changes codon GAG to GTG. Valine is a hydrophobic amino acid on the outside of the molecule. This makes sickle cell hemoglobin (HbS) less soluble than normal hemoglobin. HbS crystallizes in the deoxy state, forms small rods, and does not allow erythrocytes to deform when passing through small blood vessels. As a result local oxygen deficiency occurs in various organs. Chronic oxygen deficiency of the brain leads to learning disability. Defective erythrocytes are destroyed and cause hemolysis. Chronic anemia results in numerous sequelae such as heart failure, liver damage, and infections. In 2010, approximately 305,800 children with sickle cell disease were born, approximately two-thirds in Africa. Approximately 50% die before their fifth birthday.

C. Selective advantage for HbS heterozygotes in areas of malaria

In 1954, Allison suggested that individuals heterozygous for the sickle cell mutation are protected from severe malaria infection. Erythrocytes of heterozygotes for the sickle cell mutation are a less favorable environment for the malaria parasite than those of normal homozygotes. Heterozygotes have a higher probability of survival and of being able to reproduce. The sickle cell mutation has arisen independently at least four or five times in different malaria-infested regions.

Further Reading

Allison AC. Polymorphism and natural selection in human populations. Cold Spring Harb Symp Quant Biol 1964; 29: 137–149

Benz EJ Jr. Disorders of hemoglobin. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 852–861

Gammon K. Gene therapy: editorial control. Nature 2014; 515(7526): S11–S13

Herrick JB. Peculiar elongated and sickle-shaped red blood cell corpuscles in a rare case of severe anemia. Arch Intern Med (Chic) 1910; 6: 517–521

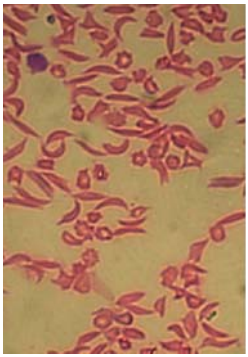
Ingram VM. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. Nature 1956; 178(4537): 792–794

Gravitz L, et al. Sickle-cell disease. Nature 2014; 515 (7526): S1–S15

Neel JV. The inheritance of sickle cell anemia. Science 1949; 110(2846): 64–66

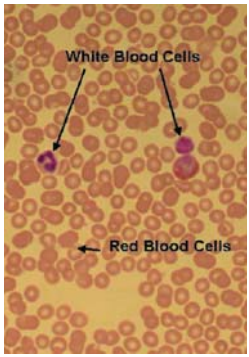
Pauling L, et al. Sickle cell anemia a molecular disease. Science 1949; 110(2865): 543–548

Rees DC, et al. Sickle-cell disease. Lancet 2010; 376 (9757): 2018–2031



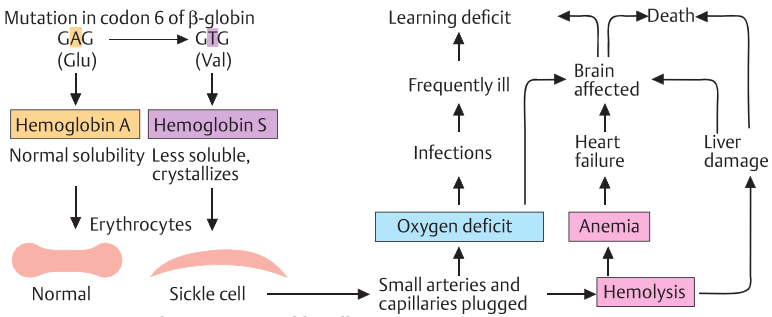
1. Sickle cells in blood smear

2. Sickle cells in blood vessels

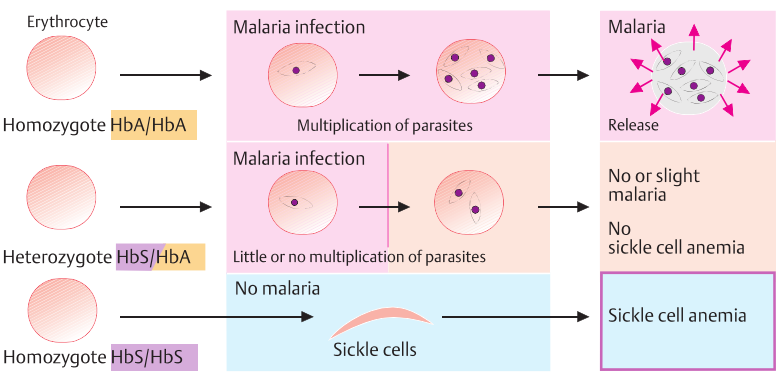


3. Normal blood smear

A. Sickle cells



B. Consequences of a mutation: sickle cell anemia



C. Selective advantage of HbS heterozygotes in areas of malaria

Mutations in Globin Genes

Mutations in globin genes cause three different types of hemoglobin disorders: (1) structural hemoglobinopathies, (2) different genetic types of thalassemia caused by impaired globin synthesis, and (3) hereditary persistence of fetal hemoglobin (HPFH). Some variants combine features of structural hemoglobinopathies and thalassemia. More than 750 human hemoglobin variants with single amino acid substitutions in one of the globin chains are known. These are distributed to the types of globin genes as follows: 217 mutations in α genes, 362 in β genes, 70 in γ genes, in 32 δ genes, 19 variants involving two mutations, 27 deletions, 6 insertions, 4 deletion/insertions, and 10 hybrid fusions.

In some cases, a globin chain is elongated or shortened or fused with parts of others. The functional consequences of mutations are decreased elasticity of the molecule, altered oxygen affinity, or instability.

A. Structural changes in the β-globin gene

More than 300 point mutations in the β-globin gene and over 100 in one of the α-globin genes have been documented. Two clinically important mutations affect codon 6: the sickle cell mutation, 6 Glu to Val (sickle cell hemoglobin, HbS, see previous page) and 6 Glu to Lys (hemoglobin C, HbC, incorporating lysine instead of glutamic acid in codon 6). Compound heterozygotes with the HbS mutation on one chromosome and the HbC mutation on the other (HbSC) show the sickle phenomenon. The marked methemoglobin formation in $Hb_{Zürich}$ and $Hb_{Saskatoon}$ results from substitutions for histidine (His) in codon 63, which alter the oxygen-binding region of the hemoglobin molecule. HbE ($\alpha 2\beta 2^{26Glu \rightarrow Lys}$) is very common in Thailand, Cambodia, and Vietnam.

B. Unequal crossing-over

The sequence homology between the globin genes may lead to nonhomologous pairing and unequal crossing-over during meiosis. A characteristic example is $Hemoglobin_{Gun\ Hill}$, described in 1968. This variant results from pairing of codon 90 with codon 95, 91 with 96, etc. As a result, codons 91 to 95 are subsequently deleted in one strand and duplicated in the other (not shown). The result is an unstable hemoglobin, of which more than 90 are known.

C. Fusion hemoglobin

Fused or hybrid hemoglobin variants probably also result from unequal crossing-over involving parts of adjacent genes. The first is Hb_{Lepore}, described in 1962. Here the first 50 to 80 amino acids of the δ chain are fused with the last 60 to 90 residues of the normal C-terminal amino acids of the β chain. The complementary situation is Hb anti-Lepore, a δβ fusion gene together with the normal δ and β genes.

D. Hemoglobins with a chain elongation

More than 10 globin chain variants with an elongated chain are known. Single base substitutions in a chain termination codon, frameshift mutations, or mutations affecting the translation initiator methionine are responsible for this type of variant. In $Hb_{Cranston}$, an insertion in codon 145 of the β chain changes UAU (tyrosine) to AGU (serine). This changes the stop codon after position 146 into ACU (tyrosine) and causes a read-through until codon 157, and consequently a chain that is 10 amino acids too long (**1**).

In HbConstant Spring (**2**), the α chain is elongated by a mutation in the stop codon UAA to CAA, which codes for glutamine (Gln).

Further Reading

Baglioni C. The fusion of two peptide chains in hemoglobin Lepore and its interpretation as a genetic deletion. Proc Natl Acad Sci U S A 1962; 48: 1880–1886

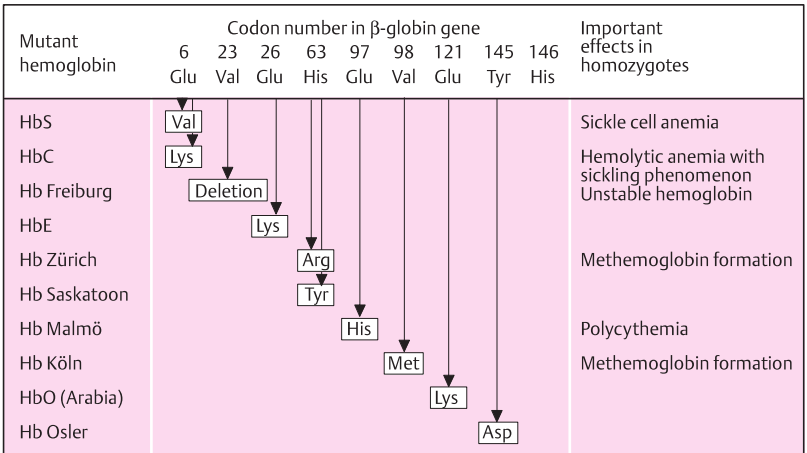
Benz EJ Jr. Disorders of hemoglobin. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 852–861

Beta-globin Mutation Database. Available at: http://globin.cse.psu.edu. Accessed May 4, 2017

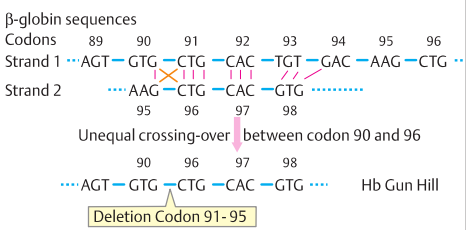
Old J. Hemoglobinopathies and thalassemias. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. Philadelphia, PA: Churchill Livingstone-Elsevier, 2013

Rieder RF, Bradley TB Jr. Hemoglobin Gun Hill: an unstable protein associated with chronic hemolysis. Blood 1968; 32(3): 355–369

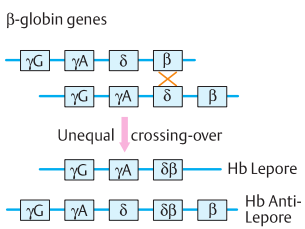
Weatherall DJ, et al. The hemoglobinopathies. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. New York, NY: McGraw-Hill, 2001: 179–190. Available at: http://www.ommbid.com. Accessed May 4, 2017



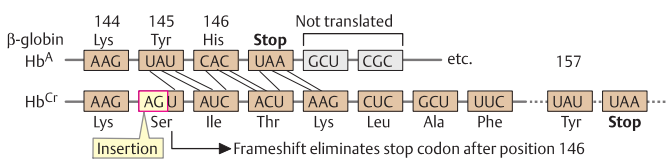
A. Structural changes in the β-globin gene



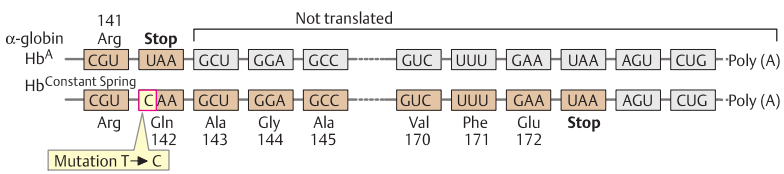
B. Unequal crossing-over



C. Fusion hemoglobin



1. Hemoglobin Cranston: chain elongation by frameshift



2. Hemoglobin Constant Spring: chain elongation by mutation in the stop codon

D. Hemoglobin with a chain elongation

The Thalassemias

Thalassemias are inherited disorders of α- or β-globin synthesis. Decreased or absent synthesis of a globin chain results in an anemic state of variable, often life-threatening severity. The thalassemias occur predominantly in regions where malaria is endemic (see p. 126). The term is derived from the Greek Θαλασσα (thalassa, sea). More than 300,000 children with a severe hemoglobin disorder are born each year, most in low- or middle-income tropical countries (Weatherall, 2010).

A. Thalassemia, a chronic anemia

Thalassemia is a chronic anemia of various grades of severity, depending on the type, first described by Cooley in 1925. It may be associated with massive extramedullary hematopoiesis (outside the bone marrow) in the liver and spleen, causing both organs to be enlarged, or bone. Infections, malnourishment, and other signs add to severe illness. Thalassemias are classified into α-thalassemias, β-thalassemias, or unstable δβ fusion globin. (Photographs from Weatherall and Clegg, 2001.)

B. β-Thalassemia and α-thalassemia

More than 200 different point mutations in the β-globin gene cause a wide spectrum of different genotypes and phenotypes in the β-thalassemias (**1**). Complete absence of the β chain (β^0) is distinguished from decreased levels of β chains. Since there are four α-globin loci, α-thalassemias (**2**) have a complex pattern of genotypes: αααα/ααα- (silent carrier); a mutation in two α loci on the same chromosome in *cis:* αα/α-α- (thal-1); or a mutation in two α loci on different chromosomes in *trans:* α-/α- (thal-2). Since the α gene loci are located within a 4-kb region of homology interrupted by small, nonhomologous regions, they are prone to mispairing and nonhomologous crossing-over during meiosis, resulting in deletion or duplication.

C. Spectrum of mutations in β-thalassemia

Mutations occur throughout the β gene. The entire gene may be deleted. Regulatory regions upstream (5′ direction) may be affected, resulting in decreased transcription. Nonsense mutations leading to a truncated globin due to a frameshift, missense mutations, and splice mutations occur. (Figure adapted from Antonarakis et al., 1985.)

D. Haplotypes in RFLP analysis

Many common mutations in the β globin gene are in linkage disequilibrium with polymorphic restriction sites in the β globin complex. Thus, certain mutations linked to a particular haplotype can be identified by restriction fragment length polymorphisms (RFLPs, see p. 58) or by SNPs (single nucleotide polymorphisms, p. 72). For example, seven polymorphic restriction sites (red arrows) define nine haplotypes, of which five are shown (A: + – – – – + +; B: – + + – + + +; etc.). Prior to modern direct DNA analysis this information was used for simplified genetic diagnosis.

α-Thalassemia may be associated with mental retardation. Two different syndromes exist: ATR-16 syndrome (OMIM 141750) associated with a large (1–2 Mb) deletion at 16pter-p13.3 including the α-globin gene cluster; the other is an X-linked disorder (OMIM 301040; gene locus at Xq13).

Further Reading

Benz EJ Jr. Disorders of hemoglobin. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 852–861

Cao A, et al. Beta-thalassemia. Genet Med 2010; 12 (2): 61–76

Higgs DR et al. Thalassaemia. Lancet 2012; 379: 373–385

Old J. Hemoglobinopathies and thalassemias. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. Philadelphia, PA: Churchill Livingstone-Elsevier, 2013

Origa R. Beta-thalassemia. Genet Med 2017; 19: 609–619

Piel FB, et al. The α-thalassemias. N Engl J Med 2014; 371(20): 1908–1916

Thein SL. The molecular basis of β-thalassemia. Cold Spring Harb Perspect Med 2013; 3(5): a011700

Weatherall DJ. The inherited diseases of hemoglobin are an emerging global health burden. Blood 2010; 115(22): 4331–4336

Weatherall DJ, Clegg JB. the Thalassaemia Syndromes. 4th ed. Oxford: Blackwell Science, 2001

Weatherall DJ, et al. The hemoglobinopathies. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. New York, NY: McGraw-Hill, 2001: 179–190. Available at: http://www.ommbid.com. Accessed May 4, 2017

Different forms of thalassemia:

- α: Decreased synthesis of α-globin
- β: Decreased synthesis of β-globin
- δβ: Decreased synthesis of δ- and β-globin

→ Unstable hemoglobin → Chronic anemia

A. Thalassemia, a chronic anemia

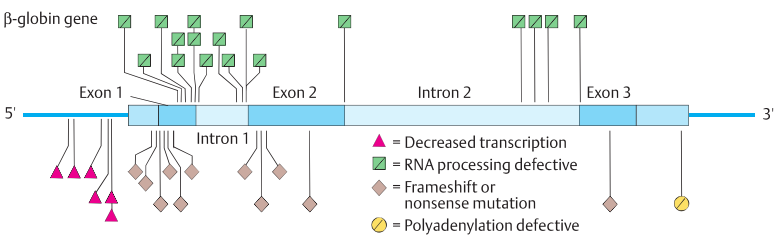
1. β-Thalassemias

Genotype		Phenotype
$\frac{+}{-}$	β^0 heterozygote	Thalassemia minor (asymptomatic)
$\frac{(+)}{(+)}$	β^+ heterozygote	
$\frac{(+)}{(+)}$	β^+ homozygote	Thalassemia intermedia (not transfusion dependent)
$\frac{+}{-}$	β^0 heterozygote	
$\frac{-}{-}$ or $\frac{(+)}{-}$	β^0 homozygote (β^0 Thalassemia)	Thalassemia major (transfusion dependent)
	β^+/β^0 homozygote (β^+ Thalassemia)	

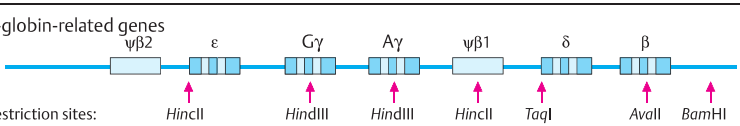
2. α-Thalassemias

Genotype	Phenotype
α α / α α	Normal
α – / α α	"Silent carrier" (normal)
– – / α α (thal-1)	Thalassemia
α – / α – (thal-2)	
α – / – –	HbH disease (HbH = β_4)
– – / – –	Hydrops fetalis

B. β-Thalassemia and α-thalassemia



C. Spectrum of mutations in β-thalassemia



Frequency	Haplotype							
47%	A	+	–	–	–	–	+	+
17%	B	–	+	+	–	+	+	+
8%	C	–	+	–	+	+	+	–
1%	D	–	+	–	+	+	–	+
12%	E	+	–	–	–	–	+	–

D. Haplotypes in RFLP analysis

Hereditary Persistence of Fetal Hemoglobin (HPFH)

Hereditary persistence of fetal hemoglobin (HPFH; OMIM 141749) is characterized by high levels of fetal hemoglobin (HbF) in adult life (5–30% instead of less than 1%). Clinically HPFH is relatively benign, although HbF is not optimally adapted to postnatal conditions. Analysis of HPFH has yielded insight into the control of globin gene transcription and the effects of mutations in noncoding sequences. In some conditions, HbF may be the only β-globin-like gene product formed.

A. Large deletions in the β-globin gene cluster

HPFH is caused by at least 25 different mutations (deletions or point mutations) within the β-globin gene cluster or point mutations in the γ-gene promoter regions. Several very large deletions in the β-globin gene cluster region are known, especially at the 3′ end. Different types of deletions occur in different ethnic populations, reflecting that they originated at different points in time. δβ-Thalassemia and failure of β-globin production have been the result in some cases.

B. Mutations in noncoding sequences of the promoter region

Hereditary persistence of fetal hemoglobin can result from mutations in noncoding sequences of the promoter region upstream (5′ direction) of the β-globin gene cluster (on the 5′ side of the γ-globin genes). Even though the highly conserved sequences CACCC, CCAAT, and ATAAA are not affected, the number of observed mutations substantiates the significance of the remaining noncoding sequences (long-range transcription control). They are probably required for the changes in transcription control in the different gene loci that occur during embryonic and fetal development. (Figure adapted from Gelehrter and Collins, 1990.)

C. Frequent mutations of β-thalassemia in different populations

β-Thalassemia mutations occur in different ethnic populations with different frequencies, as shown here for 10 examples. The difference in types and distribution of mutations allows relatively simple screening tests to identify individuals who might have an increased risk for a severe thalassemia in their offspring. According to estimates of the World Health Organization, approximately 275 million persons are heterozygous for hemoglobin diseases worldwide. Substantial numbers are due to the β-thalassemias in Asia (over 60 million), α^0-thalassemia in Asia (30 million), HbE/β-thalassemia in Asia (84 million), and sickle cell heterozygosity in Africa (50 million), India, the Caribbean, and the US (approximately 50 million). At least 200000 severely affected homozygotes are born annually: approximately 50% with sickle cell anemia and 50% with thalassemia (Weatherall, 2010).

Further Reading

Antonarakis SE, et al. DNA polymorphism and molecular pathology of the human globin gene clusters. Hum Genet 1985; 69(1): 1–14

Benz EJ Jr. Disorders of hemoglobin. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 852–861

Gelehrter TD, Collins FS. Principles of Medical Genetics. Baltimore, MD: Williams & Wilkins, 1990

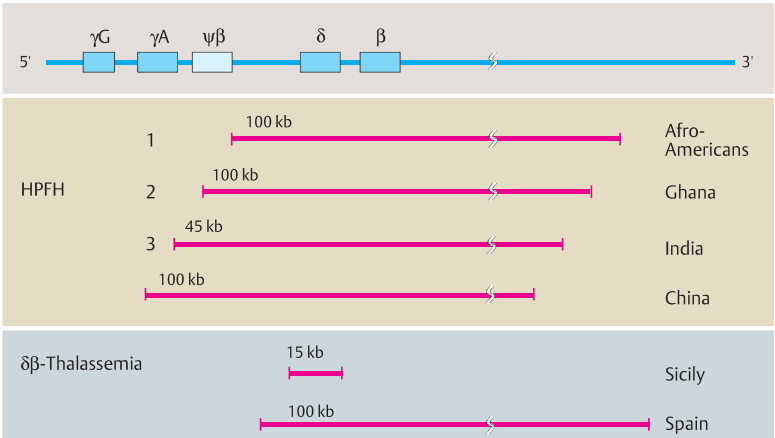
Kan YW, et al. Deletion of the beta-globin structure gene in hereditary persistence of foetal haemoglobin. Nature 1975; 258(5531): 162–163

Orkin SH, Kazazian HH jr. The mutation and polymorphism of the human β-globin gene and its surrounding DNA. Annu Rev Genet 1984; 18: 131–171

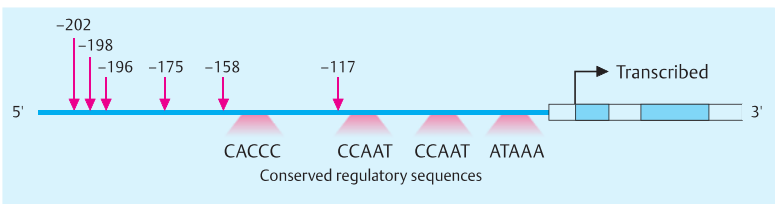
Stamatoyannopoulos G, et al., eds. The Molecular Basis of Blood Diseases. 4th ed. Philadelphia, PA: W. B. Saunders, 2001

Weatherall DJ. The inherited diseases of hemoglobin are an emerging global health burden. Blood 2010; 115(22): 4331–4336

Weatherall DJ, et al. The hemoglobinopathies. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. New York, NY: McGraw-Hill. Available at: http://www.ommbid.com. Accessed May 4, 2017



A. Large deletions in the β-globin cluster



B. Mutations in noncoding sequences in the promoter of γ-globin genes cause hereditary persistence of fetal hemoglobin (HPFH)

β-thal mutation	Frequency	Ethnic group	Type
Intron 1 (110 G→ A)	35%	Mediterranean	β^+
Codon 39 (C→T)	27%	Mediterranean	β^0
TATA-Box (-29 A→G)	39%	Afro-American	β^+
Poly A (T→ C)	26%	Afro-American	β^+
Intron 1 (5 G→C)	36%	India	β^+
Partial deletion (619 nt)	36%	India	β^0
Codon 71-72 frameshift	49%	China	β^0
Codon 71-72 frameshift	49%	China	β^0
Intron 2 (654 C→T)	38%	China	β^0
Codon 41-42 -CTTT	frequent	Southeast Asia	β^0

C. Frequent mutations of β-thalassemia in different populations

Mammalian Sex Determination

A series of early embryonic sequential binary activating or inhibitory developmental processes in mammals determine whether anatomical development will result in a male or a female. Genes expressed at specific stage of development in relevant tissues initiate a sex determination pathway toward one sex or the other. The first decision is made by the presence or absence of a functional Y chromosome. Thereafter, indifferent stages of the gonads and of the internal and external genital organs differentiate into those of the female or male gender.

A. Role of the mammalian Y chromosome

The crucial role of the Y chromosome in man became evident in 1959 from chromosomal analysis of two disorders: Klinefelter syndrome and Turner syndrome. Individuals with Klinefelter syndrome have two X chromosomes and a Y chromosome (XXY) as discovered by Jacobs and Strong in 1959, but clearly have a male phenotype, although this is incompletely developed (p. 384). Even the presence of several X chromosomes does not result in a female phenotype as long as one functional Y chromosome is present. In contrast, individuals with Turner syndrome have only one X chromosome and no Y chromosome (45,XO) discovered by Ford et al. in 1959, but nevertheless show a female phenotype, although it is incompletely developed and usually accompanied by malformations (p. 384). Thus, the Y chromosome is required to induce male development.

B. Sex-determining region SRY

In human males, a small region of approximately 35 kb on the distal short arm of the Y chromosome at Yp11.2 is required to induce male development. This region is named SRY (sex-determining region Y, OMIM 480000). SRY is located just proximal to the pseudoautosomal region 1 (PAR1) within Y-chromosomal interval 1A1. Presence of SRY results in a development toward a male and its absence toward a female.

C. *SRY* gene

The *SRY* gene consists of a single exon with 841 bp spanning 3.8 kb. It encodes a transcription factor that is a member of the high-mobility group (HMG)-box family of DNA-binding proteins. A motif TATAAA binds to transcription factor TFIID. It is transcribed into RNA of 1.1 kb. From a coding region of 612 bp, a protein of 204 amino acids is translated (**1**). The conserved HMG motif binds to DNA and causes reversible bending (**2**). The bending opens the double helix and permits access of transcription factors. HMG proteins are nonhistone DNA-binding proteins. *SRY* is part of a network of genes regulating developmental processes. Its binding target is SOX9 (608160), a transcription factor required for both sex and skeletal development. Mutations result in campomelic dysplasia with sex reversal (114290). (Photo in 2 kindly provided by Dr Michael A. Weiss, Cleveland, United States, from Li et al., 2006.)

D. Sry-transgenic male XX mouse

The crucial role of SRY was shown in transgenic mice. When a 14-kb DNA fragment containing the mouse *Sry* gene is inserted into the blastocyst of a chromosomally female (XX) mouse, a male phenotype develops. (Figure from Koopman et al., 1991.)

E. Time pattern of Sry expression

Expression of Sry is limited to short period of time during embryonic development. In a mouse embryo with XY chromosomes, Sry is expressed only between days 10.5 and 12.5 of embryonic development. (Figure from Koopman et al., 1991.)

Further Reading

Ellis PJ, Erickson RP. The sex determination pathway. In: Erickson RP, et al., eds. Epstein's Inborn Errors of Development. 3rd ed. Oxford: Oxford University Press, 2016

Graves JA. In retrospect: Twenty-five years of the sex-determining gene. Nature 2015; 528(7582): 343–344

Knower KC, et al. Failure of SOX9 regulation in 46XY disorders of sex development with SRY, SOX9 and SF1 mutations. PLoS One 2011; 6(3): e17751

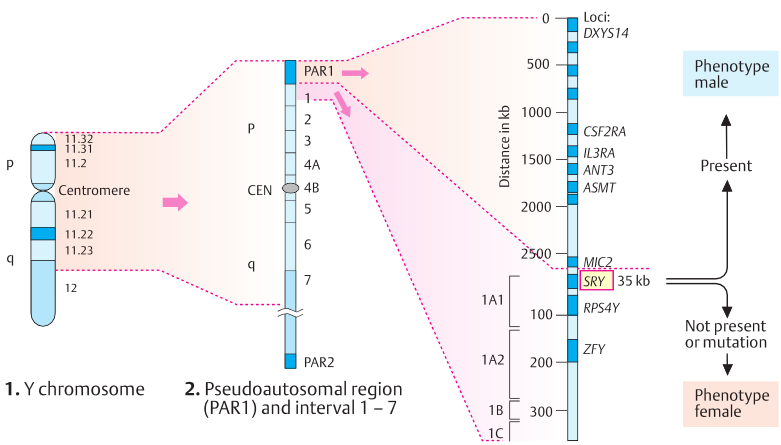
Koopman P, et al. Male development of chromosomally female mice transgenic for Sry. Nature 1991; 351(6322): 117–121

Li B, et al. SRY-directed DNA bending and human sex reversal: reassessment of a clinical mutation uncovers a global coupling between the HMG box and its tail. J Mol Biol 2006; 360(2): 310–328

Sekido R. SRY: A transcriptional activator of mammalian testis determination. Int J Biochem Cell Biol 2010; 42(3): 417–420

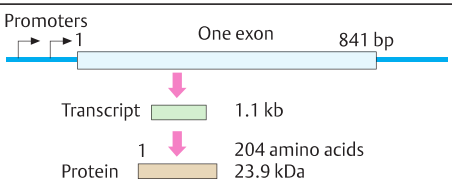
Y chromosome present		No Y chromosome present	
46, XY	Normal male	46, XX	Normal female
Additional X chromosome		One X chromosome only	
47, XXY	Klinefelter syndrome male	45, XO	Turner syndrome female

A. Role of mammalian Y chromosome



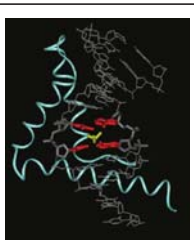
1. Y chromosome **2.** Pseudoautosomal region (PAR1) and interval 1 – 7

B. Sex-determining region *Sry* on the Y chromosome **3.** PAR1 and interval 1A – 1B

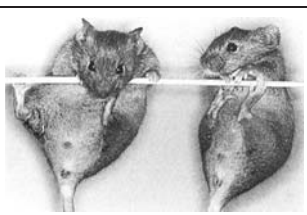


1. The *Sry* gene and protein

C. *Sry* gene

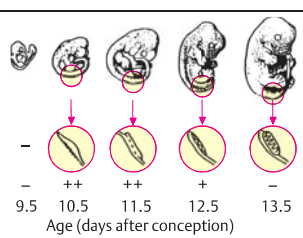


2. *SRY* binds to DNA



XY normal male XX with *Sry* gene (male)

D. *Sry*-transgenic male XX mouse



E. Time pattern of *Sry* expression

Sex Differentiation

Sex differentiation follows sex determination in a series of consecutive activating or inactivating developmental processes during early mammalian embryogenesis, resulting in either female or male gender. From initially undifferentiated anatomical structures, male or female development is induced under the influence of specific genes.

A. Gonads and external genitalia

At about the end of the sixth week of pregnancy in humans, after the primordial germ cells have migrated to the initially undifferentiated gonads, an inner portion (medulla) and an outer portion (cortex) of the gonads can be distinguished (**1**). The cortex develops into an ovary, the medulla into a testis. The early embryonic testes develop under the influence of the testis-determining factors (TDF), the *SRY* gene (in humans), or the ovary in its absence. Under the influence of estradiol, müllerian ducts develop in females and wolffian ducts develop in males. In XY embryos, testes develop at about the 10th week of pregnancy under the influence of a testis-determining factor (TDF), the *SRY* gene (OMIM 480000). If this is not present, ovaries develop.

The excretory ducts differentiate under the influence of the hormones produced by the early gonads (**2**). The early embryonic testis produces two hormones: testosterone, with a male differentiating effect; and the müllerian inhibition factor MIF (anti-müllerian hormone, AMH; OMIM 600957). The müllerian ducts, precursors of the fallopian tubes, the uterus, and the upper vagina, develop when a male differentiating influence is absent. In the male pathway, their development is suppressed by AMH. AMH shares homology with transforming growth factor beta (TGFβ, OMIM 190180). The wolffian ducts, precursors of the male efferent ducts (vas deferens, seminal vesicles, and prostate), develop under the influence of testosterone, a male steroid hormone formed in the fetal testis. If testosterone is absent or ineffective, the wolffian ducts degenerate.

The external genitalia develop after the gonads have differentiated into testes or ovaries. In humans this occurs relatively late, in the 15th to 16th week of pregnancy (**3**). Full development of male external genitalia depends on a derivative of male-inducing testosterone, 5-dihydrotestosterone, a metabolite of testosterone produced by the enzymatic action of 5α-reductase.

The differentiation of the gonadal ridge into the bipotential gonad (**4**) requires genes encoding transcription factors SF1 (OMIM 601516), WT1 (OMIM 607102), and LHX9 (OMIM 606066). The undifferentiated gonad develops into an ovary under the influence of the genes *DAX1* (OMIM 300473) on Xp21.2 and *WNT4* (OMIM 603490). The early stages of testis development require SRY and SOX9 (OMIM 608160). In addition other genes are involved (not shown).

B. Sequence of events in sex differentiation

Four levels of sex differentiation can be defined: (1) genetic, (2) gonadal, (3) anatomical (external genitalia), and (4) psychological, from early childhood on. A fifth, the legal gender, recorded as “female” or “male” in all legal documents, can be added. Each level is reached in a series of time-regulated successive steps. The male differentiating effect of testosterone depends on the function of an intracellular androgen receptor (see next page). Whereas the first three levels of differentiation are anatomically defined, the psychological may be less clearly marked.

Further Reading

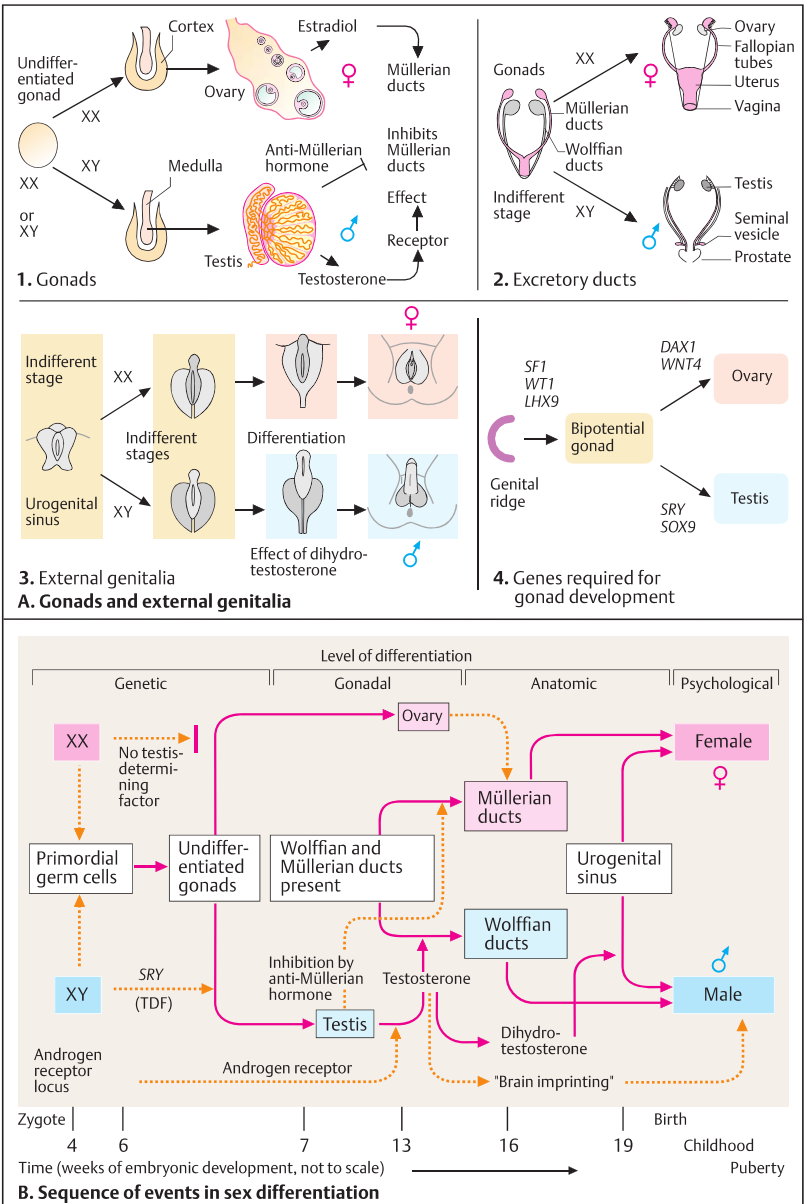
Acherman JC, et al. Disorders of sexual differentiation. In: Longo DL, et al., eds. Harrison’s Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 3046–3055

Cox JJ, et al. A SOX9 duplication and familial 46,XX developmental testicular disorder. N Engl J Med 2011; 364(1): 91–93

Ellis PJ, Erickson RP. The sex determination pathway. In: Epstein’s Inborn Errors of Development. 3rd ed. Oxford: Oxford University Press, 2016

Gilbert SF, Barresi MJF. Developmental Biology. 11th ed. Sunderland: Sinauer, 2016

Goodfellow PN, et al. SRY and primary sex-reversal syndromes. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016. Available at: www.ommbid.com. Accessed May 7, 2016


Undifferentiated gonad
XX
XY
XX
or
XY
Cortex
Ovary
Medulla
Testis
Estradiol
Müllerian ducts
Anti-Müllerian hormone
Inhibits Müllerian ducts
Effect
Receptor
Testosterone
1. Gonads
Gonads
XX
Müllerian ducts
Wolffian ducts
Indifferent stage
XY
Ovary
Fallopian tubes
Uterus
Vagina
Testis
Seminal vesicle
Prostate
2. Excretory ducts
Indifferent stage
XX
Urogenital sinus
XY
Indifferent stages
Differentiation
Effect of dihydrotestosterone
3. External genitalia
A. Gonads and external genitalia
SF1
WT1
LHX9
Genital ridge
Bipotential gonad
DAX1
WNT4
Ovary
SRY
SOX9
Testis
4. Genes required for gonad development
Level of differentiation
Genetic
Gonadal
Anatomic
Psychological
XX
No testis-determining factor
Ovary
Female
Müllerian ducts
Primordial germ cells
Undifferentiated gonads
Wolffian and Müllerian ducts present
Urogenital sinus
Wolffian ducts
XY
SRY
(TDF)
Inhibition by anti-Müllerian hormone
Testosterone
Male
Testis
Dihydrotestosterone
Androgen receptor locus
Androgen receptor
"Brain imprinting"
Zygote
Birth
4
6
7
13
16
19
Childhood
Time (weeks of embryonic development, not to scale)
Puberty
B. Sequence of events in sex differentiation

Disorders of Sexual Development

Disorders of sexual development (DSD) may occur at any level of sex determination or differentiation (see previous page). DSD can be classified according to (1) aberrations of the XY chromosome system, (2) abnormal gonadal development, (3) impaired androgen synthesis or androgen effect, (4) androgen excess, and (5) syndromic forms as part of a generalized disorder involving other tissues or organs. In some forms, the gonads do not correspond to the genital duct system or external genitalia, or the external genitalia may be ambiguous, that is, partly female and partly male (pseudohermaphroditism). In true hermaphroditism, gonads contain both testicular and ovarian tissues. Therapy requires that the underlying developmental error be correctly diagnosed.

A. XX males and XY females

These are examples of an impaired XY system. Normally, the male-determining Y-specific DNA sequences (*SRY* gene; OMIM 480000) remain on the Y chromosome during the homologous pairing and crossing-over during meiosis. Since *SRY* is located very close to the pseudoautosomal region 1 (PAR1), crossing-over outside PAR1 transfers the *SRY* region to the X chromosome and results in a male individual with an XX karyotype (XX male syndrome; OMIM 278850). The complementary Y chromosome does not receive *SRY*, which results in a female phenotype with XY chromosomes (XY female gonadal dysgenesis sex reversal OMIM 400044).

B. Point mutations in the *SRY* gene

Mutations in the human *SRY* gene lead to sex reversal, that is, failure of normal male sex differentiation in the presence of a Y chromosome. *SRY* encodes a 204-amino-acid protein containing a DNA-binding domain of 79 highly conserved amino acids, the HMG box (high-mobility group protein), in the middle section between amino acids 58 and 137. Most point mutations and small deletions cluster in this region. This results in complete or partial gonadal dysgenesis. (Figure adapted from McElreavey and Fellous, 1999.)

C. Androgen receptor disorders

Testosterone binds to the androgen receptor and enters the nucleus (**1**). The activated hormone-receptor complex (TR and DR) induces the differentiation of the wolffian ducts. Dihydrotestosterone (DHT), converted from testosterone in the urogenital sinus by 5α-reductase, also requires a receptor (DR) in order to induce the male differentiation of the urogenital sinus into the external genitalia. Mutations in the gene encoding an androgen receptor (OMIM 313700) on Xq12 result in a broad spectrum of disorders including androgen insensitivity syndrome (AIS, 300068), formerly called testicular feminization, and in spinal and bulbar muscular dystrophy type Kennedy (SBMA, 313200). XY Individuals affected with AIS have a normal male karyotype XY, a normal *SRY* gene and testes that produce testosterone. However, since testosterone cannot exert any effect, the resulting phenotype is female (**2**). Partial androgen insensitivity results in incomplete external male genitalia (Reifenstein syndrome, 312300, Mongan et al., 2015).

For classification of DSD, see appendix Table 17 (p. 415).

Further Reading

Acherman JC, et al. Disorders of sexual differentiation. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 3046–3055

Bashamboo A, McElreavy K. Mechanisms of sex determination in humans: Insights from disorders of sex development. Sex Dev 2016; 10(5–6): 313–325

Medscape. Disorders of sex development. Available at: http://emedicine.medscape.com/article/1015520

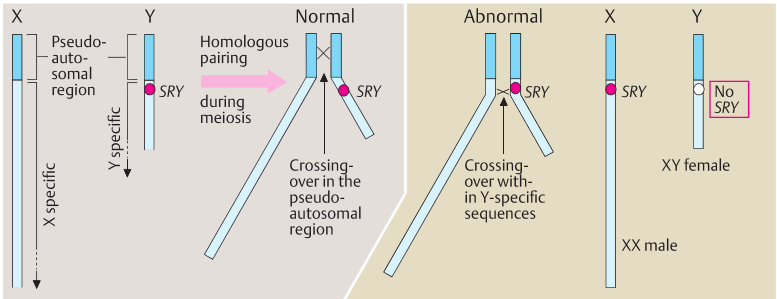
Griffin JE, et al. The androgen resistance syndromes: steroid 5 α-reductase deficiency, testicular feminization, and related disorders. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016. Available at: www.ommbid.com. Accessed May 13, 2017

McElreavey K, Fellous M. Sex determination and the Y chromosome. Am J Med Genet 1999; 89(4): 176–185

Mendonca BB, et al. 46,XY disorders of sex development (DSD). Clin Endocrinol (Oxf) 2009; 70(2): 173–187

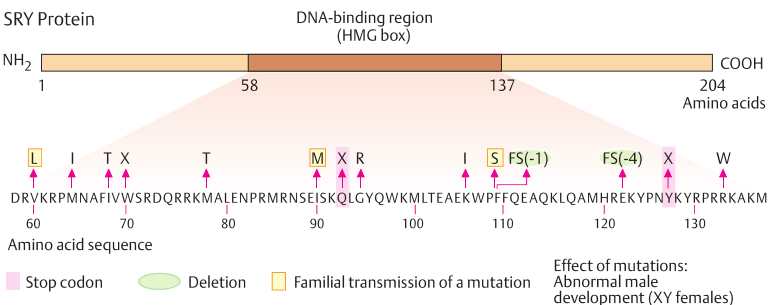
Mongan NP, et al. Androgen insensitivity syndrome. Best Pract Res Clin Endocrinol Metab 2015; 29(4): 569–580

Ohnesorg T, et al. The genetics of disorders of sex development in humans. Sex Dev 2014; 8(5): 262–272

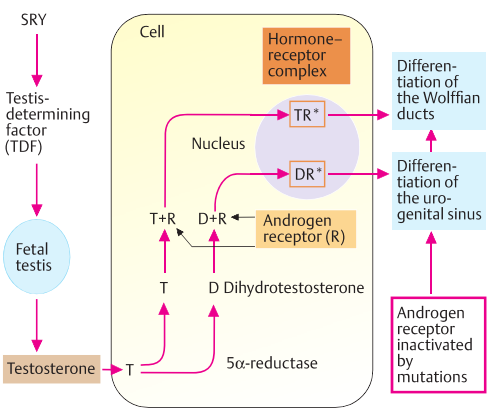


1. *SRY* remains on the Y chromosome **2.** Transfer of *SRY* to the X chromosome

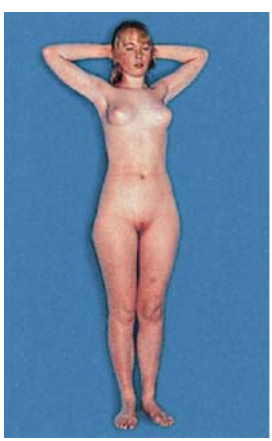
A. XX males and XY females



B. Point mutations in the *SRY* gene



1. Scheme of functional events



2. Phenotype

C. Androgen receptor and TFM

Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia (CAH), also called adrenogenital syndrome (OMIM 201910), is an autosomal recessive deficiency of cortisol resulting from insufficient biosynthesis of steroid 21-hydroxylase, a microsomal cytochrome P450 enzyme. A compensatory increase in adrenocortical hormone (ACTH) excretion causes hyperplasia of the adrenal cortex with increased prenatal production of androgenic steroids. With a frequency of approximately 1 in 5,000 newborns, it is the most common form of aberrant sex differentiation with highly variable clinical manifestations. The heterozygote frequency is 1:50 in Europe and North America.

A. Overview

The main characteristics of this disease are listed (**1**). Ambiguous or virilized genitalia are present in newborn girls (**2**). The biochemical defect results in hyperplasia of the adrenals (**3**). Inheritance is autosomal recessive (**4**). Untreated or poorly treated children show advanced growth and early puberty, but are short as adults because of early closure of the epiphyseal plates. Such girls develop a male physical appearance (**5**). CAH occurs in three major manifestations, depending in part on the type of mutation: (1) a severe salt-losing form, in approximately 60 to 75% of patients; (2) a virilizing form without salt loss; and (3) an attenuated, late-onset form. The severe salt-losing form is life-threatening in newborns.

B. Biochemical defect

The enzymatic defect blocks the conversion of progesterone to deoxycortisol (DOC) by hydroxylation at position 21 (steroid 21-hydroxylase). As a result, the plasma concentration of 17-hydroxy-progesterone is increased.

C. The *CYP21* gene structure

The *CYP21A2* gene (OMIM 613815) encodes the 21-hydroxylase enzyme. It is located at 6p21.33 in tandem with three closely linked genes upstream (in the 5′ direction): the complement 4A (*C4A*) gene (OMIM 120810), a *CYP* pseudogene (*CYP21P*) without function. These genes lie within the class III HLA genes of the major histocompatibility complex (MHC; see p. 306). The *CYP21A2* gene has 10 exons and spans nearly 6 kb. Deletions and nonsense or frameshift mutations result in the severe salt-losing form. Missense mutations occur in the salt-losing and virilizing forms. Most patients are compound heterozygotes. The sites of seven mutations are shown. Deletions account for approximately 20% of the classical salt-losing form. Duplications occur without clinical consequences.

D. Crossing-over events

The four genes, *C4A*, *CYP21P*, *C4B*, and *CYP21*, share structural and sequence similarity as a result of a duplication event in evolution. This predisposes to mispairing and unequal crossing-over at meiosis. Gene conversion occurs in approximately 75% of patients. It results from mismatch between the functional *CYP21* gene and the nonfunctional pseudogene *CYP21P*, which converts part of the *CYP21* gene to a pseudogene (*CYP21P*).

E. Molecular analysis

This example of a previous approach shows the detection of an 8-bp deletion by a semiquantitative PCR method. In normal individuals (lanes 2, 4, 6–8) the intensity of a 952-bp fragment (derived from exons 1–3) and a 200-bp fragment is about the same. In heterozygotes (lanes 5 and 9), the 952-bp fragment is less intense than in controls, and it is absent in the patient (lane 3). The fragments at the bottom of each lane represent parts of the β-actin gene for control. (Figure courtesy of Dr Alireza Baradaran, Mashhad, Iran, now Vancouver, from Vakili et al., 2005.)

Further Reading

Donohoue PA, et al. Congenital adrenal hyperplasia. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016. Available at: www.ommbid.com

Merke DP, Bornstein SR. Congenital adrenal hyperplasia. Lancet 2005; 365(9477): 2125–2136

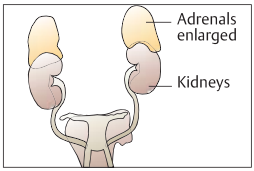
Speiser PW et al. Congenital adrenal hyperplasia. J Clin Endocrinol Metabol 2010; 95: 4133–4160

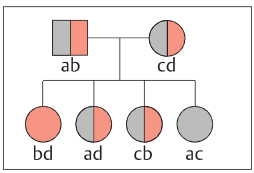
Lin-Su K, et al. Genetic disorders of the adrenal gland. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Vakili R, et al. Molecular analysis of the CYP21 gene and prenatal diagnosis in families with 21-hydroxylase deficiency in northeastern Iran. Horm Res 2005; 63(3): 119–124

1.
- Cortisol deficiency
- Salt loss
- Prenatal virilization
- Adrenal hyperplasia
- Steroid 21-hydroxylase deficiency

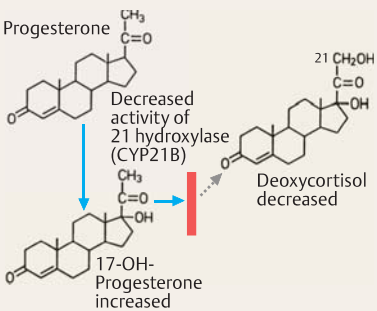
2.

3. 

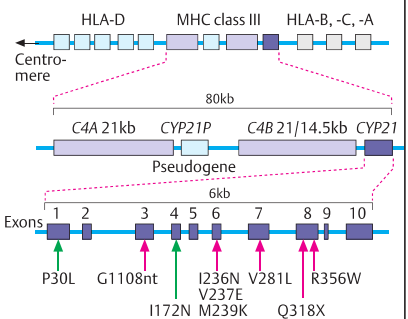
4. 

5.

A. Clinical phenotype and formal genetics

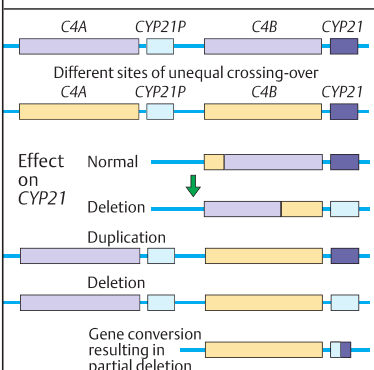


B. Biochemical defect



Selected mutations (→ salt-losing; → simple virilizing)

C. The *CYP21* gene structure

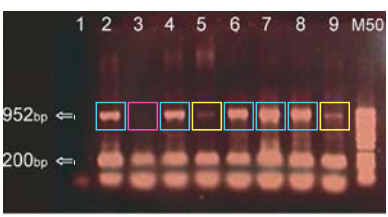


D. Crossing-over events in the *CYP21* gene

Normal: lanes 2, 4, 6, 7, 8

Homozygous deletion: lane 3

Heterozygous deletion: lanes 5, 9



E. Molecular analysis

Rhodopsin, a Photoreceptor

Vertebrates have evolved sensory systems that enable them to perceive light, sound, taste, and smell, and convert this into advantageous behavioral patterns. Two types of photoreceptors in specialized cells of the retina in the eye provide color vision and light perception in the dark. The human retina has approximately 6 million cone cells (*cones*) for color vision and approximately 110 million rod cells (*rods*) that operate in weak light. The photoreceptor responsible for weak light is *rhodopsin*. It is a G-protein-coupled receptor activated by light. Only rods contain the trimeric G protein coupled to rhodopsin.

A. Rod cells

A rod cell is a highly specialized cell containing the photoreceptor rhodopsin. An outer segment of a rod contains approximately 1,000 discs with approximately 4×10^7 molecules of rhodopsin. The discs are folded by the protein peripherin. The trimeric G protein coupled to rhodopsin is called *transducin*. At the other end of a rod is an inner segment with the cell nucleus, endoplasmic reticulum, Golgi apparatus, and mitochondria. Each rod has a synapsis. From here, a signal is transmitted to the optic nerve and from there to the visual cortex of the brain.

B. Photo excitation

In 1958, George Wald and coworkers discovered that light isomerizes 11-*cis*-retinal (**1**) into all-*trans*-retinal (**2**). This structural change in femtoseconds is so great that it triggers a reliable and reproducible nerve impulse. The absorption spectrum of rhodopsin (**3**) corresponds to the spectrum of sunlight, with an optimum at a wavelength of 500 nm. In the dark, all-*trans*-retinal is converted back into 11-*cis*-retinal. All-*trans*-retinal essentially does not exist in the dark. Although vertebrates, arthropods, and mollusks have anatomically different types of eyes, all three phyla use 11-*cis*-retinal for photoactivation.

C. Light cascade

Photoactivated rhodopsin triggers a series of enzymatic reactions, known as the light cascade. First, the photoactivated rhodopsin activates a G-protein-coupled receptor protein, transducin (G_t). Activated transducin is bound to guanosyltriphosphate (GTP; see p. 188). GTP activates cyclic guanosylmonophosphate phosphodiesterase (cGMP) by binding to its inhibitory γ subunit. This converts cGMP to GMP and closes the cGMP-gated ion channels.

D. Rhodopsin

Rhodopsin is a seven-helix transmembrane protein structure with binding sites for functionally important molecules such as transducin, rhodopsin kinase, and arrestin on the cytosolic side. The binding site for the light-absorbing 11-*cis*-retinal is lysine in position 296 of the seventh transmembrane domain.

E. cGMP as transmitter of light signals

Cyclic guanosine monophosphate (cGMP) is the second messenger in the light signal transduction system. Rod outer segments normally contain a high concentration of cGMP. Light absorption by rhodopsin activates cGMP phosphodiesterase. This hydrolyzes cGMP to 5′-GMP and decreases the cGMP concentration. The high concentration of cGMP in the dark keeps cGMP-gated cation channels open. On exposure to light, the channels are closed, the membrane becomes hyperpolarized, and a nerve impulse is triggered. (Figures adapted from Stryer, 1995, and Lodish et al., 2016.)

Medical relevance

Mutations in genes encoding the proteins of the light cascade signal pathway are frequent causes of different forms of hereditary blindness resulting from pigmentary degenerative changes to the retina (see next page).

Further Reading

Alberts B, et al. Molecular Biology of the Cell. 6th ed. New York, NY: GS Garland, 2015

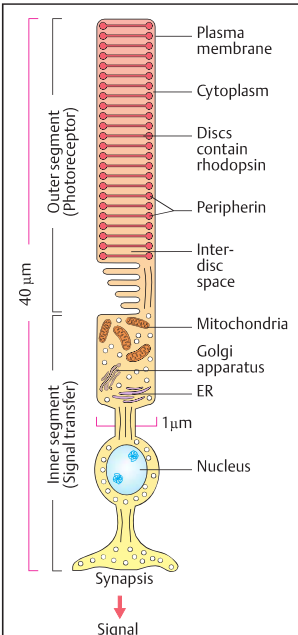
Lodish H, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman, 2016

Nathans J, Hogness DS. Isolation and nucleotide sequence of the gene encoding human rhodopsin. Proc. Nat Acad Sci 1984; 81: 4851–4855

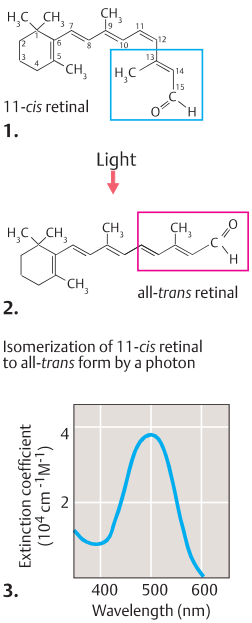
Stryer L. Molecular basis of visual excitation. Cold Spring Harb Symp Quant Biol 1988; 53(Pt 1): 283–294

Stryer L. Biochemistry. 4th ed. New York, NY: W. H. Freeman, 1995

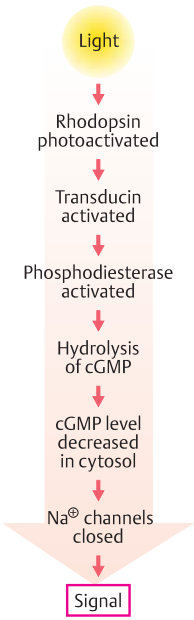
Wald G. The molecular basis of visual excitation. Nature 1968; 219(5156): 800–807



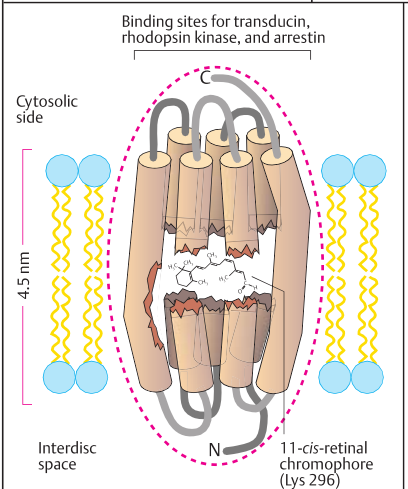
A. Rod cell



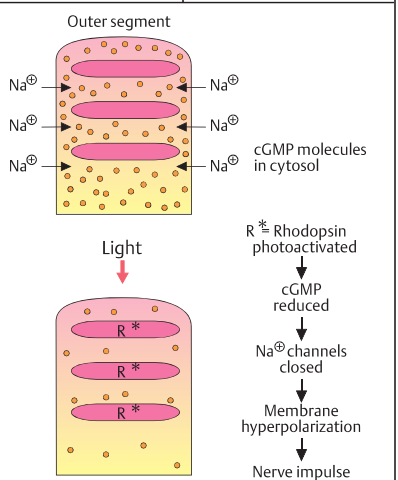
B. Photo excitation



C. Light cascade



D. Rhodopsin



E. cGMP as transmitter of light signals

Pigmentary Retinal Degeneration

Pigmentary degeneration of the retina (retinitis pigmentosa, RP) is a group of hereditary eye diseases involving more approximately 80 autosomal and 6 X-chromosomal loci (see OMIM 268000 phenotypic series) causing progressive blindness in approximately 1 in 3 to 5,000 children and adults. Mutations at these gene loci interfere with the light cascade signal pathway and cause retinal degeneration. Aside from rhodopsin, these include peripherin, cAMP phosphodiesterase, and others. Retinal degeneration may occur alone (nonsyndromic) or as part of a systemic disorder involving other organ systems (syndromic forms, see OMIM 276900, Usher syndrome).

A. Retinitis pigmentosa

Affected individuals first lose midperipheral vision; then, the visual field is progressively reduced to a small island of central vision (tunnel vision). This is associated with the inability to perceive weak light (night blindness). The fundus of the eye shows thin retinal vessels, a pale, waxy yellow optic nerve, and multiple areas of irregular hyperpigmentation and depigmentation. Approximately 30 to 40% of cases of RP are the result of autosomal dominant inheritance determined at approximately 40 gene loci. Approximately 50 to 60% of cases of RP are inherited as autosomal recessive traits at approximately 60 genes and additional gene loci (see OMIM 268000 for overview). Approximately 5 to 15% of cases of RP result from X-linked inheritance (OMIM 268000). However, in a single case without a family history, the mode of inheritance may remain uncertain. (Photograph kindly provided by Professor E. Zrenner, Tübingen, Germany.)

B. The first mutation in rhodopsin

The first mutation of rhodopsin was described by Dryja et al. in 1990. It is a transversion of cytosine (C) to adenine (A) in codon 23 of exon 1. This causes an exchange of a proline for a histidine in codon 23 (CCC for proline to CAC for histidine, designated as mutation P23H). The partial DNA sequence in the banding pattern shows the mutation as an additional band representing the adenine in codon 23. Proline in position 23 is highly conserved in evolution and occurs in more than 10 related G-protein receptors. The figure, adapted from Dryja et al., shows the previous Sanger sequence pattern, which has been replaced by automated sequencing.

C. Mutations in rhodopsin

The gene for human rhodopsin (*RHO*; OMIM 180380) has five exons and spans 5 kb. It is located on the long arm of chromosome 3 in region 2, band 2.1 (3q22.1) and encodes a transmembrane protein of 348 amino acids. Thirty-eight amino acids are invariant in vertebrates. The figure shows the distribution of selected mutations in the rhodopsin molecule. Most are missense mutations leading to an exchange of one amino acid by another. Rearrangements involving a few base pairs cause small deletions.

D. Simplified diagnosis of a mutation

The figure shows an example of indirect DNA diagnosis prior to direct mutational analysis available today. In the family shown, 13 individuals in three generations (grey squares males, grey circles females) are affected (**1**). The mutation is the same (P23H) as shown in **B**. By using allele-specific oligonucleotides and the polymerase chain reaction, individuals carrying the mutation can be distinguished from those who do not. The oligonucleotide corresponding to the mutant allele (3′-CATGAGCTTCACCGACGCA-5′) hybridizes with the mutant allele only, not with the normal allele (**2**). Individuals II-2, II-12, and III-4 were not examined. (Data from Dryja et al., 1990.)

Further Reading

Arno G et al. Mutations in REEP6 cause autosomal-recessive retinitis pigmentosa. Am J Hum Genet 2016; 99: 1305–1315

Dryja TP. Retinitis pigmentosa. Chapter 235. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016

Dryja TP, et al. A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. Nature 1990; 343(6256): 364–366

Hartong DT, et al. Retinitis pigmentosa. Lancet 2006; 368(9549): 1795–1809

Shankar SP. Hereditary retinal and choroidal dystrophies. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Sheffield VC, et al. Genomics and the eye. N Engl J Med 2011; 364(20): 1932–1942

Wright AF, et al. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. Nat Rev Genet 2010; 11(4): 273–284

A group of hereditary diseases with degeneration of photoreceptors in the retina

Night blindness

Progressive loss of vision

Frequency about 1:3500

Frequency of the different genetic forms

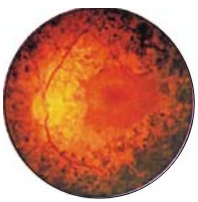
30–40% autosomal dominant
50–60% autosomal recessive
8% X-chromosomal

Important diagnostic signs

Fundus:
narrow vessels
pale optic nerve
macula changes
widened light reflex
pigment epithelium changes
electroretinogram extinguished

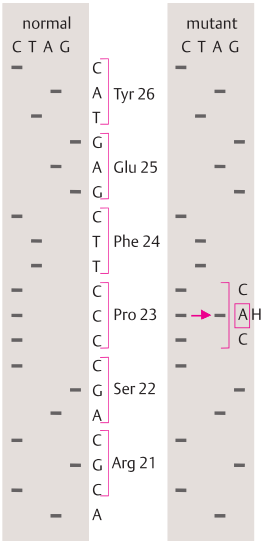
Secondary changes in the anterior chamber:
vitreous body changes

Cataract
Myopia

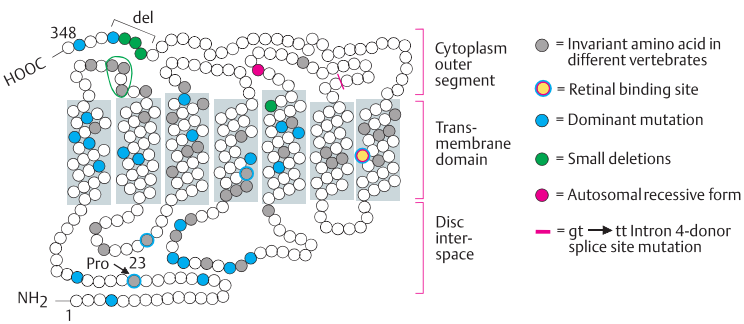


Typical fundus with pigment changes, narrow vessels, and pale, waxy optical nerve

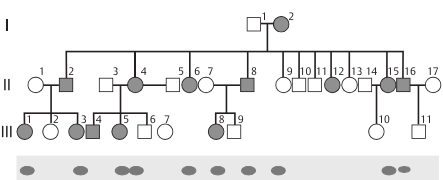
A. Retinitis pigmentosa



B. Mutation in rhodopsin



C. Mutations in rhodopsin



D. Demonstration of mutation P23H in codon 23 by oligonucleotides after PCR

Color Vision

In 1802, Thomas Young suggested that color vision in humans is trichromatic: red, green, and blue. Cone cells in the retina contain three types of photoreceptors, which respond to long-wave (red), middle-wave (green), or short-wave (blue) light. This was confirmed in 1986 when Nathans and coworkers identified the genes and molecular structure of the color receptors. X-linked red–green color vision defects in humans affect approximately 8% of males.

A. Human photoreceptors

The overlapping absorption spectra of the three classes of receptors have peaks of sensitivity at 420 nm (blue; OMIM 613522), 530 nm (green; OMIM 300821), and 560 nm (red; OMIM 300822).

B. Evolution

The genes for the visual pigment photoreceptors have arisen in evolution by duplication of an ancestral gene encoding a light-sensitive protein. Approximately 800 million years ago (mya), an ancestral visual pigment diverged by a duplication event into the rod pigment rhodopsin and another, not yet differentiated, cone pigment. A short-wavelength-responsive (blue) pigment gene and a single mid-wavelength (green–red) gene diverged approximately 500 million years ago following another duplication event. The split between New World monkeys with a single X-linked receptor and Old World monkeys with two X-linked receptors occurred approximately 30 to 40 million years ago when the single green–red pigment gene duplicated. The two copies then evolved into the green and the red pigment genes.

C. Structural similarity

Four photopigments with identical structures exist as heptahelical (seven helices) transmembrane proteins with similar sequences of amino acids. Nathans et al. in 1986 determined the percentage of identical amino acids shown in the figure in pairwise arrangements. Open circles indicate invariant amino acids present in all; dark circles indicate differences in amino acid sequences. (Figure adapted from Nathans et al., 1986.)

D. Polymorphism in the photoreceptor

Subtle differences in red light perception were detected by Motulsky and coworkers (Winderickx et al., 1992). They are due to polymorphic variants at three sites of the receptor (**1**): serine/alanine at amino acid postion 180, 230, and 233 (**2**). The distribution of red light perception showed a difference for the red/red and green mix distribution, depending on whether serine or alanine was present (**3**). (Figure adapted from Winderickx et al., 1992.)

E. Normal and abnormal red–green vision

Defects in color vision can affect any of the three photopigments.

The tandem arrangement of the X-linked genes at Xq28 encoding the red and the green receptors (**1**) predispose to color vision defects. Owing to the sequence similarity, unequal crossing-over occurs frequently in the intergenic region (approximately 15 kb), which results in deletion/duplication and different degrees of red–green vision defects (**2**). Two types of photoreceptors are involved in defective color vision: blue plus green in protanopia in the *OPN1LW* gene (OMIM 303900) or blue plus red in deuteranopia in the OPN1MW gene (OMIM 300822).

Other vision defects involve the blue pigment (tritanopia; OMIM 190900) or cause complete loss of color vision (achromatopsia; OMIM 216900).

Further Reading

Deeb SS, et al. The molecular basis of variation in human color vision. Clin Genet 2005; 67(5): 369–377

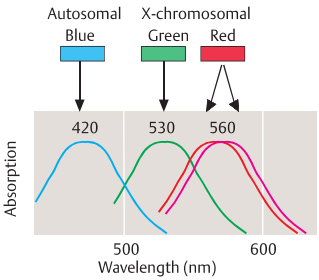
Deeb SS, et al. Color vision defects. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier

Motulsky AG, et al. Color vision and its genetic defects. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016. Available at: www.ommbid.com. Accessed May 14, 2017

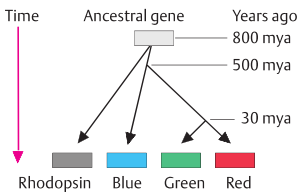
Nathans J, et al. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science 1986; 232(4747): 193–202

Winderickx J, et al. Polymorphism in red photopigment underlies variation in colour matching. Nature 1992; 356(6368): 431–433

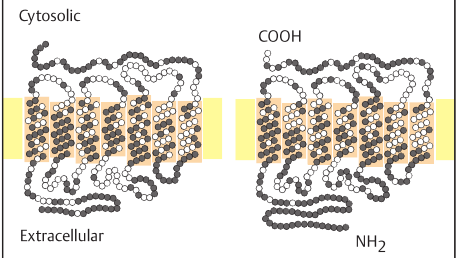
Young T. The Bakerian Lecture: On the theory of light and colours. Philos Trans R Soc Lond 1802; 92: 12–48



A. Photoreceptor proteins in rods

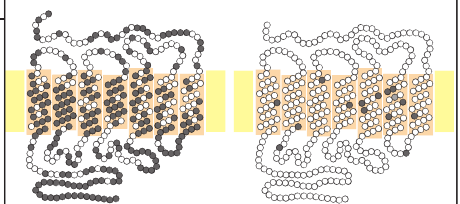


B. Evolution of genes for visual pigment photoreceptors



1. Blue/rhodopsin 75%

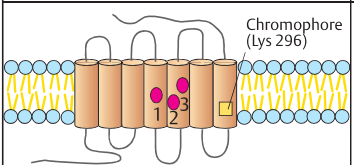
2. Green/rhodopsin 41%



3. Green/blue 44%

4. Green/red 96%

C. Similar structure of visual pigments

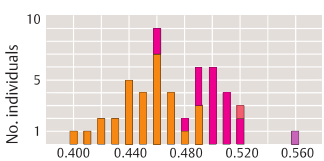


1.

Amino acid position

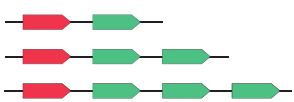
1	2	3	Frequency
180	230	233	
Ser	Ile	Ala	0.60
Ala	Ile	Ala	0.36
Ala	Thr	Ser	0.02
Ser	Thr	Ser	0.02

2.

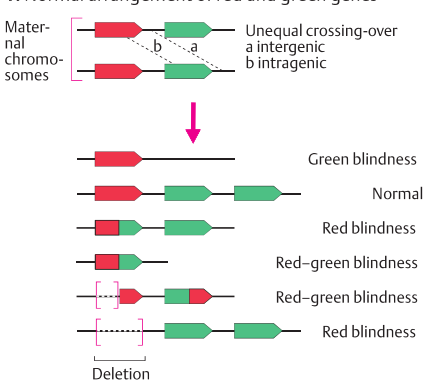


3.

D. Polymorphism in the photoreceptor for red



1. Normal arrangement of red and green genes



2. Examples of different consequences of unequal crossing-over

E. Normal and abnormal red/green vision

Auditory System

The auditory system develops and functions under the control of more than approximately 120 known genes. Mutations cause a large genetically heterogeneous group of hearing disorders ranging from congenital deafness to impaired hearing at later ages. Genetic causes account for approximately 35%, approximately 35% are acquired, and in 30% the cause remains unknown. Approximately 1 in 1,000 individuals is affected by severe hearing impairment or deafness at birth or in early childhood (prelingual deafness).

A. The main components of the ear

Sound waves are transmitted from the outer ear by vibrations in the tympanic membrane to the middle ear to the inner ear by a chain of three small movable bones (malleus, incus, and stapes). In the cochlea, the organ of Corti, the acoustic signals are amplified and processed (auditory pathway). The inner ear also harbors the vestibule with the three semicircular canals, which regulate the sense of equilibrium.

B. The cochlea

The cochlea contains three fluid-filled canals: the scala vestibuli, the scala media, and the scala tympani. Between the endolymph of the scala media and the perilymph of the scala vestibuli and scala tympani, there is an electrical potential difference of approximately +85 mV. Potassium ions (K^+) secreted from the stria vascularis of the scala media into the endolymph are recycled through supporting cells by K^+ channels and gap junctions (e.g., connexin 26 [Cx26]). The cochlea contains two types of sensory cell: one row of inner hair cells and three rows of outer hair cells. Sound-induced vibrations of the tectorial membrane deflect the stereocilia, open mechanosensitive channels, and cause an influx of K^+. The altered membrane potential elicits an impulse at the acoustic nerve, and this is transmitted to the auditory cortex of the brain.

C. Outer hair cells

The stereocilia of approximately 50,000 outer hair cells (**1**) are arranged in arrays resembling an organ pipe. Their tips are connected by tip links containing myosin and the cell adhesion molecule cadherin-23. The regular arrangement of the outer hair cells (**2**, top) is disrupted by mutations affecting the cytoskeleton (**2**, bottom).

D. Congenital deafness

The more than 120 genes involved in hereditary deafness are localized on nearly every human chromosome. Approximately 70 to 80% of mutations are autosomal recessive (gene loci designated DFNB1–DFNB105 in 2017, see 85 entries in OMIM 220290, phenotypic series), 15 to 20% autosomal dominant (DFNA1–DFNA67, see 54 entries phenotypic series in OMIM 606346, 166800), and 1 to 2% X-chromosomal (DFN1–DFN6, OMIM 304500). The more than 60 identified genes encode a variety of proteins of the acoustic pathway (see Appendix, Table 18 p. 416). The most frequent mutations are in connexin 26 (Cx26/GJB2; OMIM 121011: DFNA3, DFN; B1). Individuals carrying certain mtDNA mutations are highly sensitive to antibiotics of the aminoglycoside class (e.g., streptomycin). Seven genes (*OTC*) are involved in otosclerosis (see OMIM 166800, phenotypic series).

In 30% of the prelingual forms, other organ systems are involved in addition to the ear (syndromic forms). Usher syndrome with both deafness and retinal pigment degeneration consists of 13 different genetic types (OMIM 276900). Approximately 95% of deaf children are born to parents with normal hearing.

Further Reading

Falk RE, Pandya A. Hereditary hearing impairment. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier

Hereditary Hearing Loss. Available at: http://hereditaryhearingloss.org/. Accessed May 18, 2017

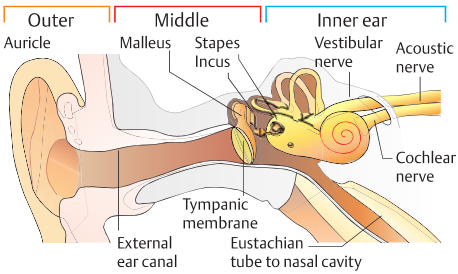
Hilgert N, et al. Function and expression pattern of nonsyndromic deafness genes. Curr Mol Med 2009; 9(5): 546–564

Lalwani AK. Disorders of hearing loss. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill Medical, 2012: 248–255

Petit C, et al. Hereditary hearing loss. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016. Available at: www.ommbid.com. Accessed May 18, 2017

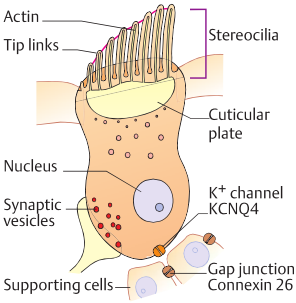
Smith RJH, et al. Sensorineural hearing loss in children. Lancet 2005; 365(9462): 879–890

Toriello H, Smith DS, eds. Hereditary Hearing Loss and Its Syndromes. 3rd ed. Oxford: Oxford University Press, 2013

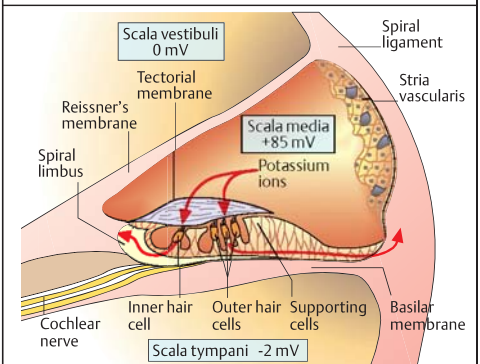


1. Cross-section of the human ear

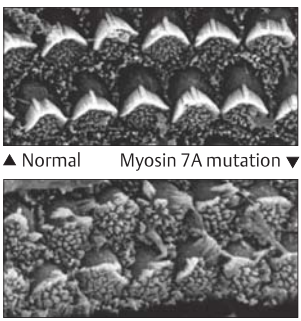
A. The main components of the ear



1. Scheme of an outer hair cell

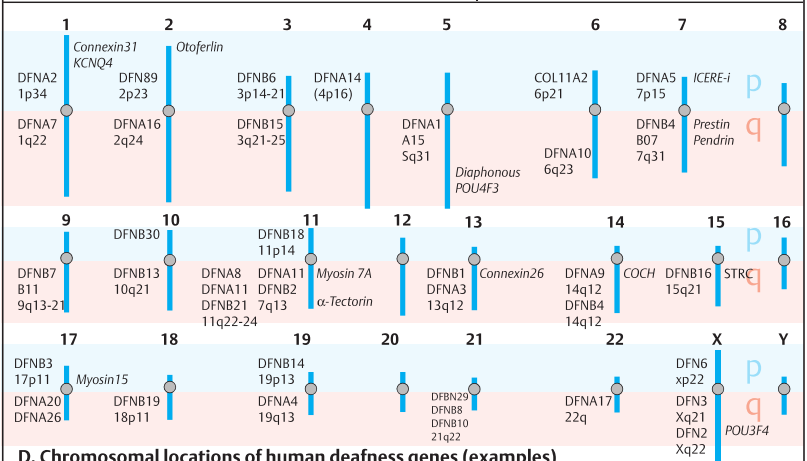


B. The cochlea



2. Stereocilia of outer hair cells in mice

C. Outer hair cells



D. Chromosomal locations of human deafness genes (examples)

Odorant Receptors

Vertebrates are able to differentiate thousands of individual odors by means of specific receptors on the cilia of olfactory neurons (odorant receptors, ORs). OR genes have arisen in evolution by many duplication events. Genes of the OR family form the largest family of genes known in mammals, and account for approximately 3 to 4% of all genes. The mammalian genome contains approximately 1,000 OR genes, and the fish genome contains approximately 100. The rat genome contains 1,866 ORs in 113 locations, consisting of 65 multigene clusters and 44 single genes. In humans, approximately 60% of the genes are pseudogenes without function.

A. Sensory olfactory nerve cells

The peripheral olfactory neuroepithelium of the nasal mucous membrane consists of three cell types: olfactory sensory neurons with axons leading to the olfactory bulb, supporting cells, and basal cells. The latter serve as stem cells that replace olfactory sensory neurons. Each olfactory neuron is bipolar, with olfactory cilia in the lumen of the nasal mucous membrane and a projection to the olfactory bulb. From there, odorant-induced signals are transmitted via the olfactory nerve to the brain.

B. Odorant-specific receptor

The odorant-specific receptor is a GTP-binding protein (see p. 188) with a specific stimulatory α-subunit, the G_{olf}. Binding of the odorant ligand to the receptor activates G_{olf}, which in turn activates adenylate cyclase. The increase in cAMP (cyclic 3′,5′-adenosine monophosphate) opens a cAMP-gated ion channel, which depolarizes the cell membrane and elicits a nerve signal. Each receptor in the cilia of the olfactory neurons binds specifically to one odorant ligand only.

C. The olfactory receptor protein

The odorant receptor is a typical seven transmembrane G-protein-coupled protein. Unlike rhodopsin, the OR proteins contain many variable amino acids, especially in the fourth and fifth transmembrane domains, and this is likely to be related to their function.

D. Exclusive gene expression

Only one allele of an OR gene is expressed and each gene is expressed in a few olfactory neurons only. Receptor-specific probes recognize only very few neurons in the olfactory epithelium of the catfish (*Ictalurus punctatus*): probe 202 hybridizes to two neurons (two black dots, **1**); probe 32 hybridizes to one neuron (**2**). Odors are distinguished in the brain according to which neurons are stimulated. (Figure adapted from Ngai et al., 1993.)

E. Subfamilies within the OR family

Olfactory receptor genes form a large family of related genes. Amino acid sequences derived from partial nucleotide sequences of cDNA clones (F2–F24) are variable, especially in transmembrane domains III and IV (**1**). For example, families F12 and F13 differ in only 4 of 44 positions. On the other hand, within a subfamily evolutionary homology is evident by considerable amino acid sequence identity (**2**). (Figures adapted from Buck and Axel, 1991, and Ngai et al., 1993.)

Medical relevance

Olfactory dysfunction affects approximately 1% of the population below the age of 60 years. Anosmia (lack of ability to smell) is associated with hypogonadotropic hypogonadism resulting in the genetically heterogeneous Kallmann syndrome involving approximately 19 autosmal loci and 1 X-chromosomal at Xp22.31 (OMIM 308700; 147950; 244200).

Further Reading

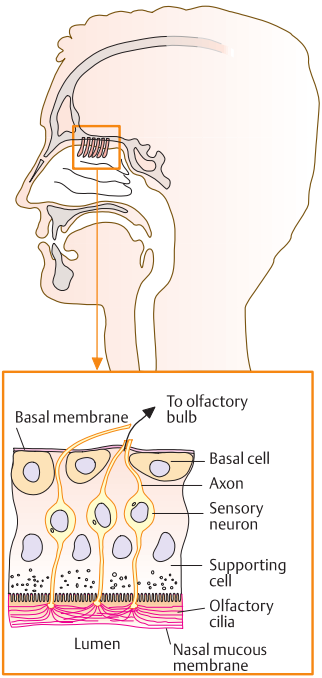
Buck L, Axel R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell 1991; 65(1): 175–187

Doty RL, Bromley SM. Disorders of smell and taste. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 241–247

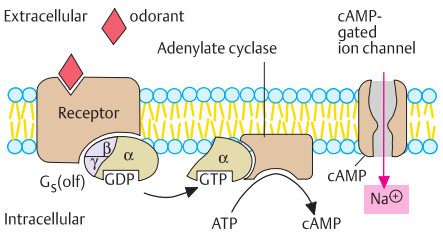
Emes RD, et al. Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents. Genome Res 2004; 14(4): 591–602

Keller A, et al. Genetic variation in a human odorant receptor alters odour perception. Nature 2007; 449 (7161): 468–472

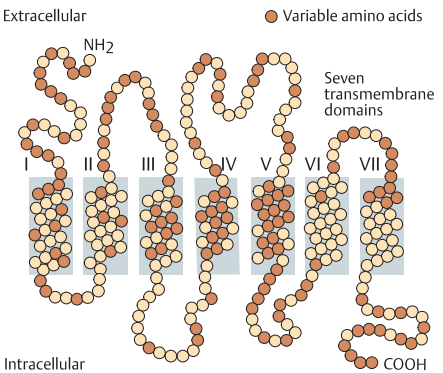
Ngai J, et al. The family of genes encoding odorant receptors in the channel catfish. Cell 1993; 72(5): 657–666



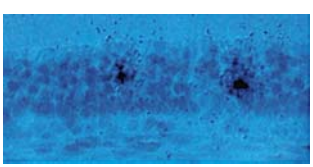
A. Olfactory nerve cells in the nasal mucous membrane



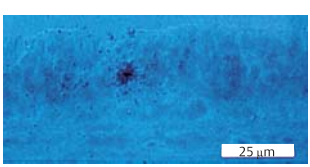
B. Odorant-specific transmembrane receptors



C. The olfactory receptor protein



1. Receptor probe 202



2. Receptor probe 32

D. Exclusive expression of OR genes

cDNA clones	Transmembrane domain						
F2	RVNE	VV I FI VVS L F LVLPFAL I IMSYV	RIVSS I LKVPSSQG I YK
F3	FLND	L V I YF T LVL L ATVPLAG I FYSYF	KIVSS I CA I SSVHGKYK
F5	HLNE	LM I L T EGAVVMVTPFVC I L I SY I	HITCAVLRVSSPRGGWK
F6	QVVE	L V S FG I AF CV I LGSCG I TLVSYA	YIITT I I K I PSARGRHR
F7	HVNE	L V I FVMGG I I L VI PFVL I I VSYV	RIVSS I LKVPSARG I RK
F8	FPSH	L TMHL V PV I L AA I SL SG I LYSYF	KIVSS I RSMSSVQGKYK
F12	FPSH	L I MNL V PVML AA I SF SG I LYSYF	KIVSS I HS I STVQGKYK
F13	FPSH	L I MNL V PVML AA I SF SG I LYSYF	KIVSS I RSVSSVKGKYK
F23	FLND	V I MYF A LVL L AVVPLL G I LYSYS	KIVSS I RA I STVQGKYK
F24	HE I E	M I I LVL AA F NL I SSL L VVL VSYL	F IL I A I LRMNSAEGRRK

1. Variable amino acid sequences

F12	FPSH	L I MNL V PVML AA I I SF SG I LYSYF	KIVSS I HS I STVQGKYK
F13	FPSH	L I MNL V PVML AA I I SF SG I LYSYF	KIVSS I RSVSSVKGKYK
F8	FPSH	LTMHL V PV I L AA I I SL SG I LYSYF	KIVSS I RSMSSVQGKYK
I12	FPSH	L I MNL V PVML GA I I SL SG I LYSYF	KIVSSVRS I SSVQGKHK
F23	FLND	V I MYF A LVL L AVVVPLL G I LYSYS	KIVSS I RA I STVQGKYK
F3	FLND	LV I YF T LVL L ATVVPLA G I FYSYF	KIVSS I CA I SSVHGKYK

2. Homology within subfamilies

E. Subfamilies within the multigene family

Mammalian Taste Receptors

Humans can distinguish five types of taste: sweet, sour, salt, bitter, and glutamate (umami, the taste of monosodium glutamate, present in Asian food). Hydrogen ions (H^+) of acids are responsible for sour taste. The salty taste is the result of direct influx of sodium ions (Na^+) from water-soluble salts. Bitter, sweet, and glutamate taste perception is mediated by families of G-protein-coupled receptor (GPCR) signaling pathway systems.

A. Mammalian chemosensory epithelia

The oral and nasal cavities of mammals contain: (1) the main olfactory epithelium (MOE) containing sensory cells with odorant receptors in the nose (see p. 378); (2) the taste sensory epithelium of the taste buds of the tongue, soft palate, and epiglottis; and (3) the vomeronasal organ (VOM, also called Jacobson's organ). The main olfactory bulb (MOB) relays signals from the MOE to the olfactory cortex of the brain. The accessory olfactory bulb (AOB) relays signals from the VOM to areas of the amygdala and hypothalamus.

B. Chemosensory signal transduction

Aside from the olfactory system (see previous page), the sensory taste (gustducin) system mediates taste perception. In addition, the vomeronasal system detects signals arriving as pheromones. Each neuron of the main olfactory sensory system (**1**) sends an axon to specific glomeruli (mitral cells), and from there to the main olfactory bulb (MOB) and the olfactory nerve. The odorant receptor (OR) gene family comprises approximately 1,000 members with distinct odorant specificity (G-olfactory proteins, G_{olf}). The bitter taste sensory system (**2**) connects axonal projections of receptor cells in the taste sensory epithelium of the taste buds to gustatory nuclei in the brain stem. Bitter taste, especially, is perceived at very low concentrations. Two families of taste receptors exist, type 1, sweet (TAS1R2 and TAS1R3, OMIM 606226, formerly T1Rs) and type 2, bitter (TAS2Rs, OMIM 604791, formerly T2Rs). They are encoded by 2 and 50 to 80 genes, respectively. A heterotrimer of TAS1R1 (606225) and TAS1R3 (605865) mediates umami taste perception.

The mammalian pheromone receptors (VN1R1 and VN2R2, OMIM 605234), located in the vomeronasal organ, form two families. They are encoded by 30 to 50 genes and over 100 genes, respectively (**3**).

C. Taste receptor gene family

The taste receptor gene family consists of intronless genes. A small family of three G-protein-coupled receptors (T1R1, T1R2, and T1R3) mediate sweet and glutamate (umami) taste perception. Bitter perception is mediated by T2R receptors. Here marked amino acid sequence differences (light fields) between the second (TM2) and third (TM3) transmembrane domains of the predicted amino acid sequences are visible for 23 different T2R receptors of human (h), rat (r), and mouse (m) origin. In contrast, conserved sequences (dark blue) indicate identity in at least half of the aligned sequences; light blue represents conserved substitutions; and the remaining are divergent regions. The T2R genes cluster on a few chromosomes: human chromosomes 5, 7, and 12, and mouse chromosomes 6 and 15.

D. Expression pattern

Unlike receptor cells of the olfactory system, individual taste receptor cells express multiple receptors (T2Rs). Up to 10 T2R probes hybridize to just a few cells, shown as darkened areas (**1**). Double-label fluorescence in-situ hybridization shows that different receptor genes are expressed in the same taste receptor cell, shown in green for T2R-7 (**2**) and red for T2R-3 (**3**). The T2Rs confer high sensitivity for bitter substances at low concentrations. (Figures adapted from Adler et al., 2000, and Dulac, 2000.)

Further Reading

Adler E, et al. A novel family of mammalian taste receptors. Cell 2000; 100(6): 693–702

Bachmanov AA, et al. Taste receptor genes. Annu Rev Nutr 2007; 27: 389–414

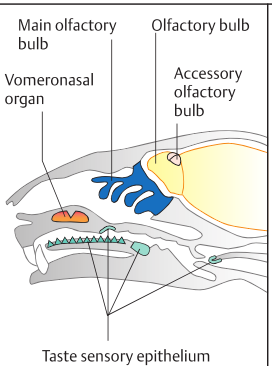
Buck LB. The molecular architecture of odor and pheromone sensing in mammals. Cell 2000; 100 (6): 611–618

Chandrashekar J, et al. T2Rs function as bitter taste receptors. Cell 2000; 100(6): 703–711

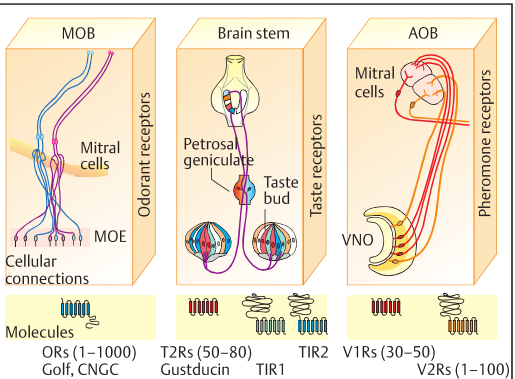
Doty RL, Bromley SM. Disorders of smell and taste. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 241–247

Dulac C. The physiology of taste, vintage 2000. Cell 2000; 100(6): 607–610

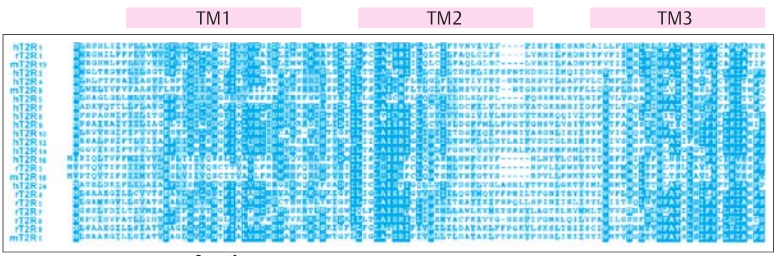
Fischer A, et al. Evolution of bitter taste receptors in humans and apes. Mol Biol Evol 2005; 22(3): 432–436



A. Mammalian chemosensory epithelia

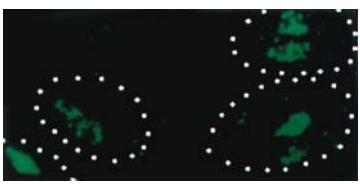


B. Chemosensory signal transduction systems

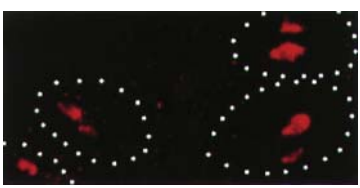


C. Taste receptor gene family

1. Ten T2R probes



2. T2R-7



3. T2R-3

D. Expression pattern of taste receptor genes

Numerical Chromosomal Aberrations

Three autosomal trisomy syndromes and three types of numerical aberrations of the X and the Y chromosome occur in liveborn human infants with an overall incidence of approximately 1:400. Trisomy arises during meiosis I or II (see p. 94) by nondisjunction as a *prezygotic* event. It may, less commonly, occur after fertilization as a *postzygotic* event in the early embryo during a somatic cell division (mitosis). In this case, the additional chromosome is present only in a certain proportion of cells. This condition is called *chromosomal mosaicism*.

A. Autosomal trisomies in humans

Trisomies for chromosome 21, chromosome 18, and chromosome 13 occur. Their incidence differs as shown and is influenced by the age of the mother. Trisomy for all other chromosomes is not compatible with postnatal life after birth and results in spontaneous abortion in the first or second trimester. The reason for this difference is the relatively low density of gene loci on chromosomes 21, 18, and 13.

B. Additional X or Y chromosome

An additional X or Y chromosome occurs with a higher incidence than autosomal trisomies and considerably less effects on the phenotype, except for the presence of a single X chromosome (monosomy X, see next page)

C. Phenotypes of autosomal trisomies in man

The three autosomal trisomies are associated with a distinctive phenotype and occur with different frequencies. Each has a distinct pattern of congenital malformations associated with variable degrees of mental impairment in trisomy 21 (Down syndrome; OMIM 190685) and complete lack of mental development in trisomies 18 and 13. Only trisomy 21 is compatible with survival into adulthood, although the overall life expectancy is approximately half of that of the normal population. The chromosomal region responsible for the phenotype Down syndrome is 21q22. Trisomy 21 may occur in sibs due to a translocation involving a chromosome 21, most often by a centric fusion (robertsonian translocation) with another acrocentric chromosome. Numerous additional manifestations (transient neonatal thrombocytosis, macrocytosis or myeloproliferative disorder, immune deficiency) and congenital malformations (congenital heart defects; duodenal stenosis; intestinal aganglionosis) may be present in Down syndrome. Always present is reduced muscular tone in newborn and infants. Children with Down syndrome grow up well integrated into family life with usually delayed developmental milestones. Males with Down syndrome are infertile, but few females have carried in infant to term.

D. Nondisjunction as a cause

All three human trisomies occur more frequently with advanced maternal age (**1**). The age of the father has no or very little influence. The faulty distribution due to nondisjunction may occur in meiosis I or II (**2**). If nondisjunction occurs in meiosis I, the three chromosomes will be different (1 + 1 + 1), whereas if nondisjunction occurs during meiosis II, two of the three chromosomes will be identical (2 + 1). In humans, approximately 70% of nondisjunctions occur in meiosis I, and 30% occur in meiosis II. In trisomy 21, the nondisjunction occurs in 90% in maternal meiosis (in meiosis I in 73%, in meiosis II in 25%). The figures for trisomy 18 and 13 are similar.

(For additional information on chromosomal aberrations, see Table 19 in the Appendix p. 417.)

Further Reading

Boué A, et al. Cytogenetics of pregnancy wastage. Adv Hum Genet 1985; 14: 1–57

Curry CJ. Autosomal Trisomies. In: Rimoin DL, et al., eds. Emery and Rimoins's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

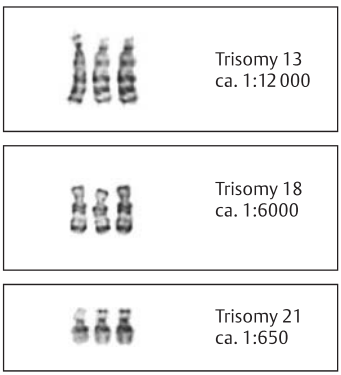
Epstein CJ. Down syndrome (trisomy 21). In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016. Available at: www.ommbid.com)

Mikhail FM. Chromosomal basis of inheritance. In: Rimoin DL, et al., eds. Emery and Rimoins's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

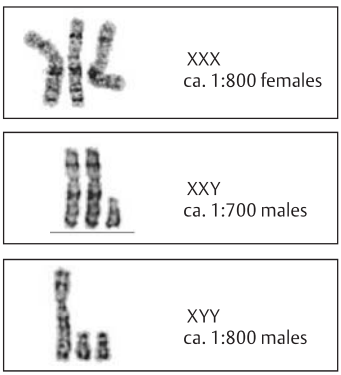
Miller OJ, Therman E. Human Chromosomes. 4th ed. New York, NY: Springer-Verlag, 2001

Roizen NJ, et al. Down's syndrome. Lancet 2003; 361 (9365): 1281–1289

Schwartz S, Hassold T. Chromosome disorders. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. Longo DL, et al., eds. New York: McGraw-Hill, 2012: 509–518



A. Autosomal trisomies in man



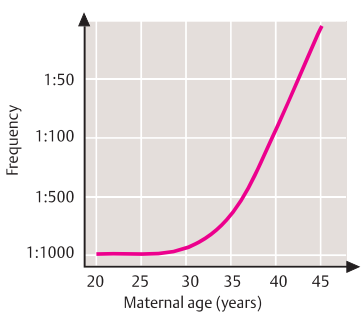
B. Additional X or Y chromosome

1. Trisomy 21

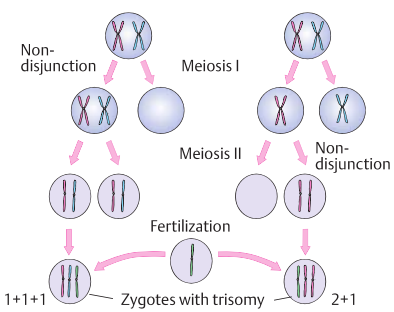
2. Trisomy 18

3. Trisomy 13

C. Autosomal trisomies



1. Trisomy 21 in liveborn infants



2. Faulty distribution of one chromosome

D. Nondisjunction as cause of trisomy

Triploidy, Monosomy X, Additional X or Y Chromosome

Other conditions associated with an abnormal number of chromosomes involve either an entire additional set of chromosomes (triploidy or tetraploidy) or the X chromosome or Y chromosome. Deviations from the normal number of X or Y chromosome comprise approximately half of all chromosomal aberrations in man (total frequency approximately 1:400).

A. Triploidy

In triploidy, either one paternal and two maternal sets (karyotype 69,XXY or 69,XXX) or one maternal and two paternal sets are present (69,XXX, 69,XYY or 69,XXY). Triploidy is associated with severe developmental failure and congenital malformations (**1**), which are not compatible with postnatal life after birth. The fetus shows numerous severe malformations (**2**), such as cardiac defects, cleft lip and palate, skeletal defects, and others. Triploidy (**3**) accounts for approximately 17% of spontaneous abortions. The causes include a diploid spermatocyte, a diploid oocyte, or fertilization of an egg cell by two spermatozoa (dispermy).

B. Monosomy X (Turner syndrome)

Monosomy X (karyotype 45,X, often written as XO) causes Turner syndrome with a very wide phenotypic spectrum, ranging from severe to very mild. The phenotype during fetal life is usually massive lymphedema of the head and neck, and large multilocular thin-walled lymphatic cysts (**1**). Congenital cardiovascular defects, especially involving the aorta, and kidney malformations are frequent. Important is fetal degeneration of the ovaries into connective tissue, as streak gonads. Small stature is always present, with an average adult height of approximately 150 cm. In many patients, the manifestations are mild (**2**), but in others, webbing of the neck (pterygium colli) may be present as a residue of the fetal lymphedema (**3**). Most patients have chromosomal mosaicism 45,X/46, XX, that is, some of their cells have normal chromosomal complements, a deletion Xp, or an isochromosome for the long arm (i[Xq]). The loss of genes on the short arm of the X chromosome (Xp) is responsible for the phenotype (see *SHOX* genes; OMIM 312865). Monosomy X is frequent, approximately 5% at conception. However, of 40 zygotes with monosomy X, only 1 will develop to birth.

C. Additional X or Y chromosome

An additional X chromosome in males (47, XXY) causes Klinefelter syndrome after puberty if untreated (**1**). This includes tall stature, absent or decreased development of male secondary sex characteristics, and infertility due to absent spermatogenesis. Testosterone substitution beginning at puberty is necessary. In contrast, an additional Y chromosome (47, XYY) does not result in a recognizable phenotype (**2**). Girls with three X chromosomes (47, XXX) are physically unremarkable (**3**). However, learning disorders and delayed speech development have been observed in some of these children more frequently than expected in girl with two X chromosomes.

D. Chromosomal aberrations in human fetuses

A wide spectrum of trisomies (and monosomies) occur at conception and lead to spontaneous abortion during the second and third months of pregnancy. The most frequent is trisomy 16, which accounts for approximately 5% of all autosomal trisomies. (Data from Lauritsen, 1982.)

Further Reading

Acherman JC, et al. Disorders of sex development. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 3046–3055

Bondy CA. New issues in the diagnosis and management of Turner syndrome. Rev Endocr Metab Disord 2005; 6(4): 269–280

DeGrouchy J, Turleau C. Clinical Atlas of Human Chromosomes. 2nd ed. New York, NY: John Wiley & Sons, 1984

Boué A, et al. Cytogenetics of pregnancy wastage. Adv Hum Genet 1985; 14: 1–57

Gravholt CH. Sex chromosome abnormalities. In: Rimoin DL, et al., eds. Emery and Rimoins's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Lauritsen JG. The cytogenetics of spontaneous abortion. Res Reprod 1982; 14: 3–4

Miller OJ, et al. Human Chromosomes. 4th ed. Heidelberg: Springer-Verlag, 2001

Otter M, et al. Triple X syndrome: a review of the literature. Eur J Hum Genet 2010; 18(3): 265–271

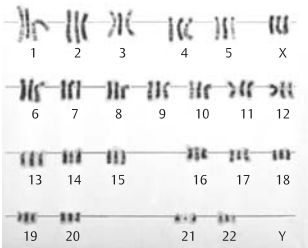
Ranke MB, Saenger P. Turner's syndrome. Lancet 2001; 358(9278): 309–314

Triploidy

- Most frequent chromosomal aberration (15%) in fetuses following spontaneous abortion
- Severe growth retardation, early lethality
- Occasional liveborn infant with severe malformation
- Dispermia a frequent cause

1.

2.



3.

A. Triploidy

1.

2.

3.

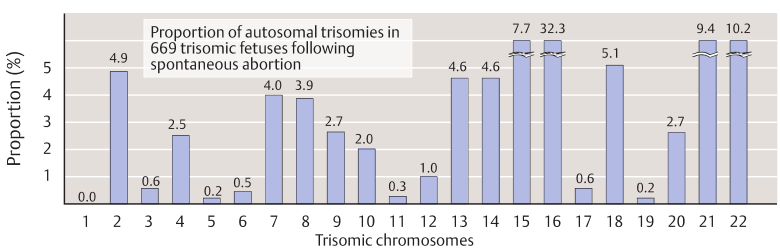
B. Monosomy X (Turner syndrome; 45,XO)

1. XXY

2. XYY

3. XXX

C. Additional X or Y chromosome



D. Wide spectrum of chromosomal aberrations in human fetuses

Microdeletion Syndromes

Microdeletion syndromes are structural chromosomal disorders that are not detectable by standard metaphase karyotyping because the size of the deletion is below the resolution of the light microscope (below 4 Mb). Microdeletions are genomic disorders (p. 418.). Most deletions occur de novo, but transmission from a parent with a balanced translocation is not rare.

Currently, approximately 60 microdeletion and microduplication syndromes have been recognized (Appendix, Table 20 p. 418). A few examples are illustrated below.

A. Deletion 5p-: cri-du-chat syndrome

This is not a microdeletion syndrome in the strict sense, because the deletion is recognizable by standard metaphase analysis in approximately 80% of patients. In 1963, Lejeune and coworkers in Paris described children with a partial deletion of the short arm of a chromosome 5 (5p-) and a characteristic pattern of dysmorphic facial features associated with impaired mental development (OMIM 123450). The name is derived from a prolonged high-pitched cry of affected infants, which resembles that of a kitten. In approximately 12% of children with cri-du-chat syndrome, one of the parents has a translocation involving a chromosome 5.

B. Deletion 4p-: Wolf–Hirschhorn syndrome

This characteristic phenotype (OMIM 194190) results from a partial deletion of variable size of the short arm of a chromosome 4. It was described in 1964 independently by U. Wolf and K. Hirschhorn and their coworkers. Mental and statomotoric retardation is associated with characteristic facial features (**1**, **2**), midline defects (cleft palate, hypospadias), coloboma of the iris, congenital heart defects, and other malformations. In approximately 80% of cases, the deletion at 4p16.3 is recognizable by light microscopy. In approximately 10 to 15% of cases, one parent carries a balanced translocation, which predisposes to deletion or duplication of this region. (Photos in A and B courtesy of Drs Beate Albrecht and Alma Küchler, Essen, Germany.)

C. Deletion 22q11

Several partially overlapping phenotypes result from deletions in region 1, band 1.2 of the long arm of chromosome 22 (22q11.2) (**1**). The most common are the DiGeorge syndrome (OMIM 188400; p. 310) and the velocardial facial syndrome (Shprintzen syndrome; OMIM 192430) with dysmorphic facial features (**2**, **3**). Approximately 80% patients have a large 1.5- to 3.0-Mb hemizygous deletion of 22q11.2 involving approximately 45 gene loci in this region, leading to phenotypic differences. (Photos in 1 and 2, courtesy of Dr Diana Mitter, Leipzig, Germany.)

D. Deletion 7q11.23

This deletion is usually associated with a recognizable phenotype, the Williams–Beuren syndrome (WBS; OMIM 194050). Facial and behavioral features associated with supravalvular aortic stenosis are its hallmark. It results from deletions of variable sizes of 1.5 to 1.8 Mb, involving approximately 28 gene loci.

E. Microdeletion syndrome 1p36

This is a relatively frequent (1:5,000) microdeletion syndrome of the subtelomeric region 1p36 (OMIM 607872) (**1**). Mental retardation associated with a prominent forehead, midfacial hypoplasia (**2**, **3**), microcephaly, muscular hypotonia, and congenital heart defects are among the phenotypic features. The diagnosis can be confirmed by array techniques, here by SNP array analysis (deletion at 1p36, arrow). (1–3, Courtesy of Prof DagmarWieczorek, Düsseldorf, 4, Dr Ludger Klein-Hitpass, Essen, Germany.)

Further Reading

D'Angelo CS, et al. Extending the phenotype of monosomy 1p36 syndrome and mapping of a critical region for obesity and hyperphagia. Am J Med Genet A 2010; 152A(1): 102–110

Spinner NB, et al. Deletions and other structural abnormalities of autosomes. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

Vissers LE, et al. Microdeletion and microduplication syndromes. Methods Mol Biol 2012; 838: 29–75

1. 5 months old

2. $3\frac{1}{4}$ years old

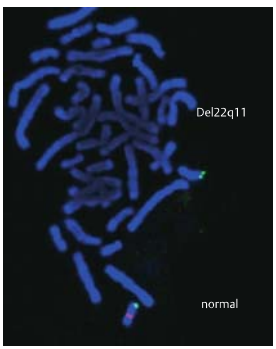
A. Deletion 5p–(Cri–du–chat syndrome)

1. $1\frac{1}{4}$ years old

2. 4 years old

B. Deletion 4p–(Wolf–Hirschhorn syndrome)

1.

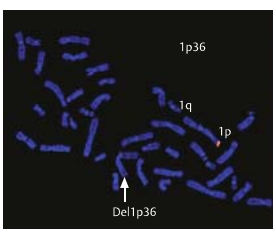


C. Deletion 22q11

2.

3.

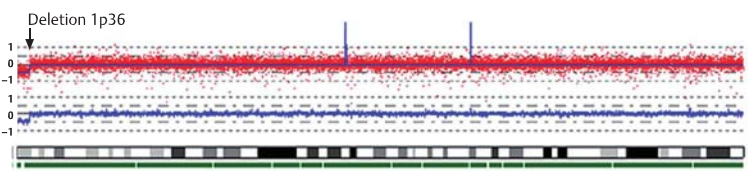
D. Deletion 7q11.23 (Williams–Beuren syndrome)



1. Fluorescence-*in-situ*-hybridization

2. 2 months old

3. 2 years old



E. Deletion 1p36

4. SNP-array (Affymetrix 259) analysis of chromosome 1

A Brief Guide to Genetic Diagnosis

The diagnosis of a genetic disease requires a systematic, interdisciplinary approach that takes many clinical and genetic considerations into account. It begins with an analysis of the phenotype. This is defined at different levels, that is, recognizable features, age of onset and progression of a disorder, laboratory and X-ray data, etc. Appropriate online systems need to be consulted. For mendelian disorders, the foremost important source of information is *Mendelian Inheritance in Man* (McKusick, 1998), freely available online as OMIM (see p. 392).

A. Genetic diagnosis, a multistep procedure

A genetic diagnosis is established in a sequence of steps, each with a binary decision. The first decision determines whether a pattern of the manifestations can be recognized. If this is the case, the next decision deals with the presumptive category of the disease considered. Although difficult to establish in practice, this decision is the basis for subsequent steps. Furthermore, as emphasized by Lupski et al. in 2011, the underlying causes of genetic diseases represent a continuum, and the assignment to a given category should not be taken too strictly. Since a particular phenotype may be caused by mutations at different loci (*locus heterogeneity*) or by different mutant alleles at the same locus (*allele heterogeneity*), caution has to be exercised not to overlook genetic heterogeneity. All genetic diagnostic procedures should be preceded by genetic counseling, which properly includes obtaining (informed) consent from the persons involved.

B. Genotype analysis by PCR typing

The polymerase chain reaction (PCR) has greatly simplified DNA-based diagnosis in cases where sequencing is not available or suitable. The figure shows how two alleles can be distinguished by the presence of a variant in one allele, but not the other. Here one allele (allele 2) contains a cleavage site and is represented by two fragments of 7 and 6 kb. The other allele (allele 1) consists of one 13-kb fragment.

Following amplification, the three possible combinations of two alleles can be visualized as the three genotypes: 1–1, 1–2, and 2–2. Each close relative of the affected individual can thus be genotyped.

C. Protein truncation test (PTT)

This is a test for frameshift, splice, or nonsense mutations that lead to a truncated protein as a result of an early stop codon created downstream of the mutation. The truncated protein is detected by an in vitro translation system. The translation will be interrupted at a premature stop codon resulting from the mutation. The size of the newly translated protein is determined by gel electrophoresis. PTT is useful in studying genes with frequent nonsense mutations, such as the *APC*, *BRCA1*, and *BRCA2* genes.

Further Reading

Aase JM. Diagnostic Dysmorphology. New York, NY: Plenum, 1990

Childs B. A logic of disease. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill. 2016. Available at: www.ommbid.com

Harper PS. Practical Genetic Counselling. 7th ed. London: Edward Arnold, 2010

Horaitis R, et al. Mutation databases: overview and catalogues. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016. Available at: www.ommbid.com

Jones KL, et al. Smith's Recognizable Patterns of Human Malformation. 7th ed. Philadelphia, PA: W. B. Saunders, 2016

Jorde LB, et al. Medical Genetics. 5th ed. Philadelphia, PA: Mosby-Elsevier, 2015

Lupski JR, et al. Clan genomics and the complex architecture of human disease. Cell 2011; 147(1): 32–43

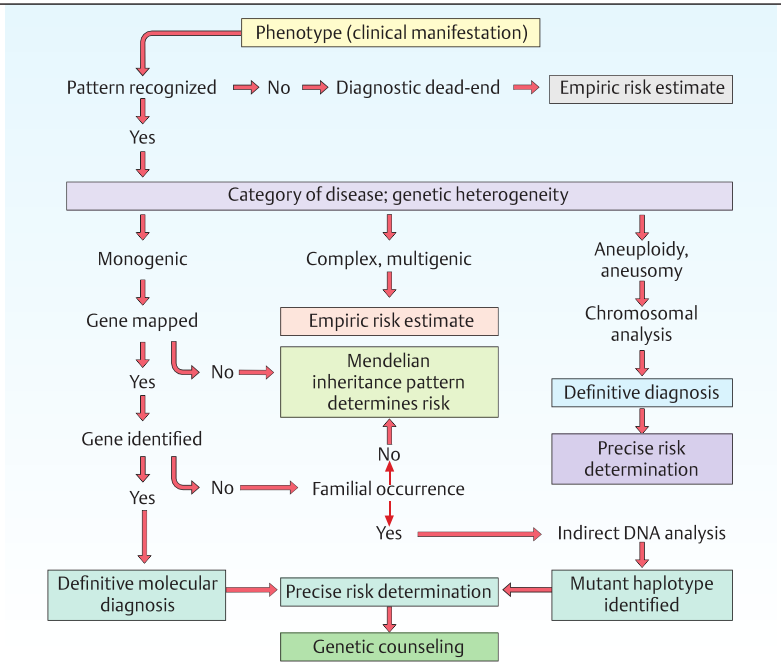
McKusick VA. Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders. 12th ed. Baltimore, MD: Johns Hopkins University Press, 1998

Misfeldt S, et al. The practice of genetics in clinical medicine. In: Longo DL, et al., eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 519–525

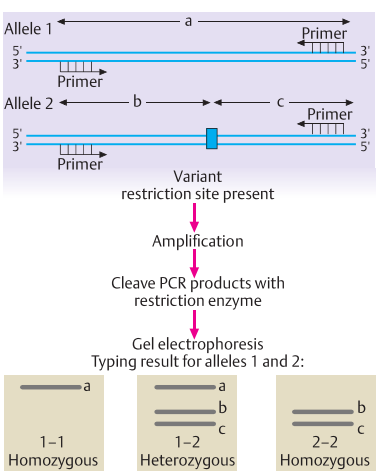
OMIM (Online Mendelian Inheritance in Man). Available at: www.ncbi.nlm.nih.gov/omim. Links to diagnostic laboratories available at: www.ncbi.nlm.nih.gov/sites/GeneTests/

Rimoin DL, et al. Emery and Rimoins's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2013

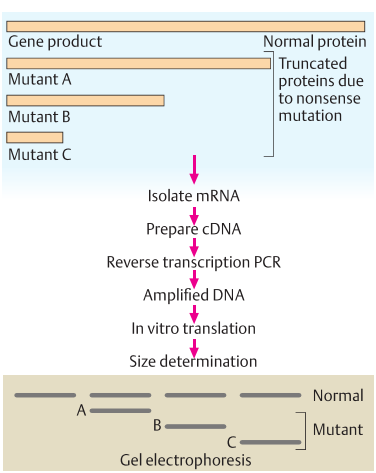
Stevenson RE, et al., eds. Human Malformations and Related Anomalies. 3rd ed. Oxford: Oxford University Press, 2015



A. Genetic diagnosis, a multistep procedure



B. Genotype analysis by PCR typing



C. Protein truncation test (scheme)

Gene and Stem Cell Therapy

Therapy of genetic diseases, a future goal, could rest on restoring function of a mutated gene, replacing it with a functional one, or inactivating it. This will require transfer of nucleic acid as the active agent, aside from other molecules or proteins (somatic gene therapy). An additional strategy is to replace lacking stem cells for tissue repair. Considerable technical difficulties and side effects have to be overcome, and ethical considerations have to be taken into account. Numerous systematic clinical trials explore the possibilities of developing a therapy of genetic diseases worldwide.

A. Principle of gene therapy

Somatic gene therapy requires a transport system to deliver correcting DNA to the somatic cells targeted. Two basic forms of somatic gene therapy are *ex vivo* and *in vivo*. In the first, DNA or a gene is introduced into cells outside the body. These are subsequently reintroduced into tissues that need to be corrected. In the second, a gene is introduced directly by a viral or a nonviral vector. The advantage of viral vectors is the relative ease with which they can enter the cells of the recipient. However, controlling viral production, carrying capacity, dependence on cell proliferation, and other aspects make application difficult.

An example for an *ex vivo* approach for the hematopoietic system is illustrated in the figure. From whole blood (**1**), the red blood cells and white blood cells are separated, and the red blood cells reinfused (**2**). From the white blood cells (**3**), immune competent CD34 cells are separated (**4**), placed together with a virus vector carrying the desired normal gene (**5**), and propagated together in a cell culture (**6**). Here the defective cells incorporate the therapeutic viral vector. These cells are reintroduced into the recipient (**7**). (Figure based on Kessler, 2012.)

B. Stem cells

Stem cells are undifferentiated progenitor cells from which differentiated cells derive. *Totipotent* stem cells can develop into a complete embryo. *Pluripotent* stem cells can form tissues derived from the endoderm, mesoderm, and ectoderm germinal layers. Stem cells divide symmetrically into two identical stem cells (self-renewal). They form a pool of cells from which progenitor cells of differentiated cells develop following an asymmetric cell division. Such cells lose the ability to undergo cell division and self-renewal.

C. Stem cell therapy

Stem cell therapy would provide the recipient with a permanent supply of cells with the ability to restore damaged tissues that lack the ability of repair. It would be especially important in organs, such as the bone marrow (hematopoietic system) and epithelial cell systems (e.g., in the gastrointestinal tract), where cells are constantly lost and replaced. Embryonic stem (ES) cells have the capacity to generate any cell type in the body; however, ES cells have the potential to form teratomas and their use is ethically controversial. Induced pluripotent stem cells (iPS cells) are somatic cells that are reprogrammed into the pluripotent state by means of four genes encoding specific transcription factors (Oct4, Sox2, Klf4, cMyc). The umbilical cord blood contains progenitor cells of the hematopoietic system, but there are limitations for usage. (Figure adapted from Nabel, 2004.)

Further Reading

Genetics Home Reference (NIH). Gene therapy. Available at: https://ghr.nlm.nih.gov/primer/therapy/. Accessed May 29, 2017

Kessler JA. Applications of stem cell biology in clincial medicine. In: Longo DL et al, eds. Harrison's Principles of Internatal Medicine. 18th ed. New York: McGraw-Hill 2012, p. 543–551

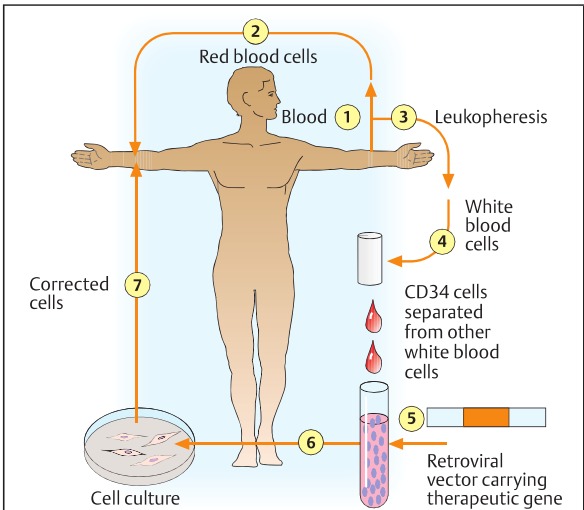
Mosteiro L, et al. Tissue damage and senescence provide critical signals for cellular reprogramming in vivo. Science 2016; 354(6315): 1020

Nabel GJ. Genetic, cellular and immune approaches to disease therapy: past and future. Nat Med 2004; 10(2): 135–141

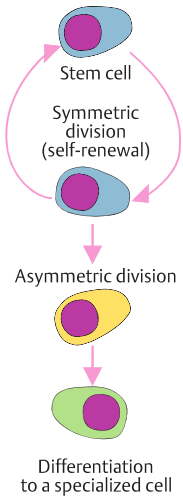
Robinton DA, et al. The promise of induced pluripotent stem cells in research and therapy. Nature 2012; 481(7381): 295–305

Sauderson NRR, et al. Gene therapy: from theoretical potential to clinical implementation. In: Rimoin DL, et al., eds. Emery and Rimoin's Principles and Practice of Medical Genetics. 6th ed. New York, NY: Elsevier, 2016

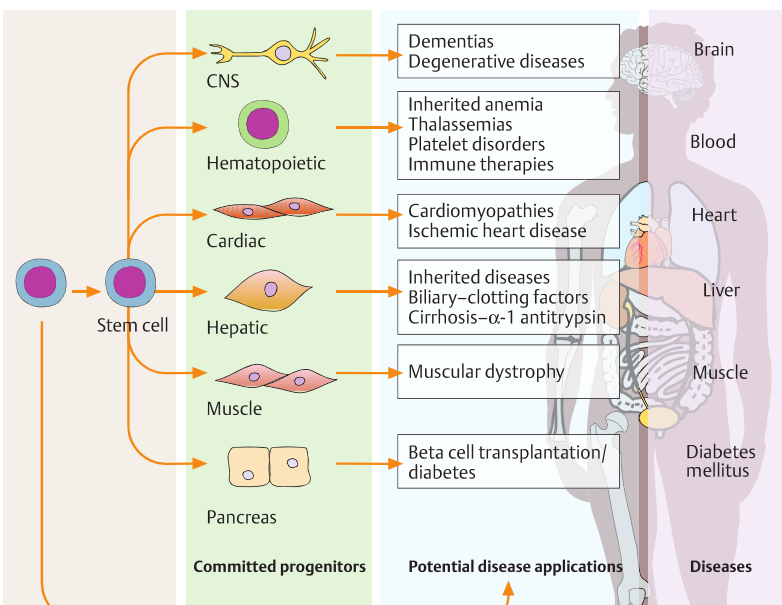
Treacy EP. Treatment of genetic disease. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill, 2016. Available at: http://ommbid.mhmedical.com. Accessed May 30, 2017



A. Principle of gene therapy



B. Stem cells



C. Prospects for stem cell therapy

Chromosomal Locations of Human Genetic Diseases

Nowhere is the completed DNA sequence of the whole human genome and each of the 24 chromosomes more apparent than in the precise map of gene loci along each chromosome. When subjected to disease-causing mutations or rearrangements, they represent a morbid anatomy of man. Today, it can be considered equivalent to Andreas Vesalius (1514–1564) seven volumes of *De Humani Corporis Fabrica* (1543) and the causal analysis of diseases *De Sedibus et Causis Morborum per Anatomen Indagatis* (1761) by Giovanni Morgagni (1682–1771) as emphasized by Victor A McKusick (1921–2008) at Johns Hopkins University in his *Mendelian Inheritance of Man: A Catalog of Human Genes and Genetic Disorders*, founded in 1966.

In this systematic collection of information about human genes and phenotypes, documented in 12 published editions between 1966 and 1998, most information about human genes and the effects of disease-causing changes is contained.

The McKusick catalog provides a systematic basis for the genetics of man comparable to the first periodic table of chemical elements by Dimitri I. Mendeleyev in 1869 or to the *Chronologisch-thematisches Verzeichnis sämtlicher Tonwerke Wolfgang Amadé Mozarts* by Ludwig Alois Ferdinand Köchel in 1862.

McKusick's first edition in 1966 contained a total of 1,487 entries. The second edition (1,545 entries) in 1968 included the first autosomal gene to be mapped. The subsequent editions reveal an entry-doubling time of approximately 15 years (3,368 in the 6th edition (1983), 5,710 entries in the 10th edition (1992), and 8,587 entries in the 12th edition (1998). In 1987, the McKusick catalog became internationally available online as OMIM from the National Library of Medicine. McKusick referred to this disease gene map as *The Morbid Anatomy of the Human Genome* (McKusick, 1998; Amberger et al., 2001). This map of disease-related gene loci assigned to specific chromosomal sites first appeared in the third edition in 1971 on a single page.

Regularly updated, OMIM is a major source of information on human genes and genetic diseases with a total of 24,065 entries as of May 26, 2017 (see the following table, OMIM Statistics). Each entry has a unique six-digit identifying number and is assigned to one of six catalogs according to genetic criteria: (1) autosomal dominant inheritance, (2) autosomal recessive, (3) X-chromosomal, (4) Y-chromosomal, and (5) mitochondrial. Autosomal entries initiated in 1994 begin with the digit 6.

The following five plates provide a limited overview of selected examples of monogenic disorders located on each human chromosome. The variable gene density and the many pseudogenes are not shown. The whole information is no longer accessible in print, but only online at Online Mendelian Inheritance in Man (OMIM).

Further Reading

Amberger JS, et al. The morbid anatomy of the human genome. In: Valle D, et al., eds. The Online Metabolic and Molecular Bases of Inherited Disease. 9th ed. New York, NY: McGraw-Hill. Available at: http://ommbid.mhmedical.com

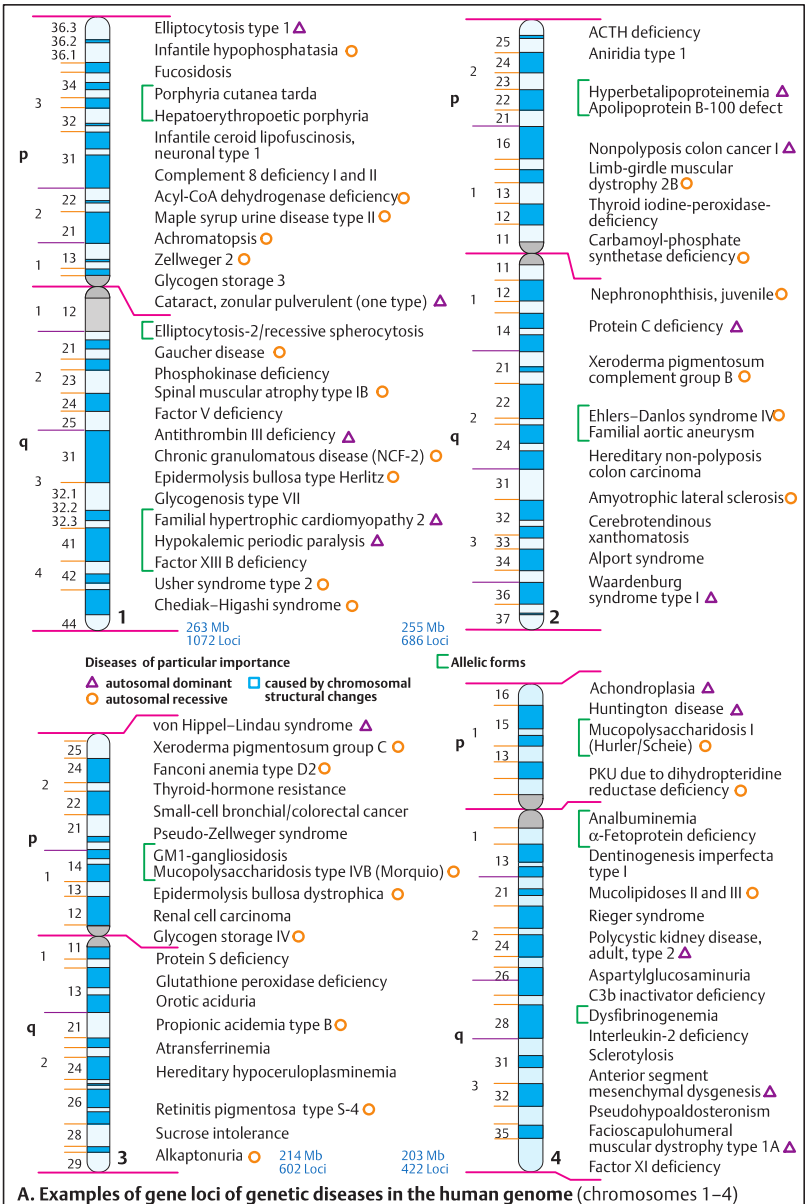
Childs B. Genetic Medicine. Baltimore, MD: Johns Hopkins University Press, 1999

McKusick VA. Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders. 12th ed. Baltimore, MD: Johns Hopkins University Press, 1998. Available at: www.ncbi.nlm.nih.gov/omim

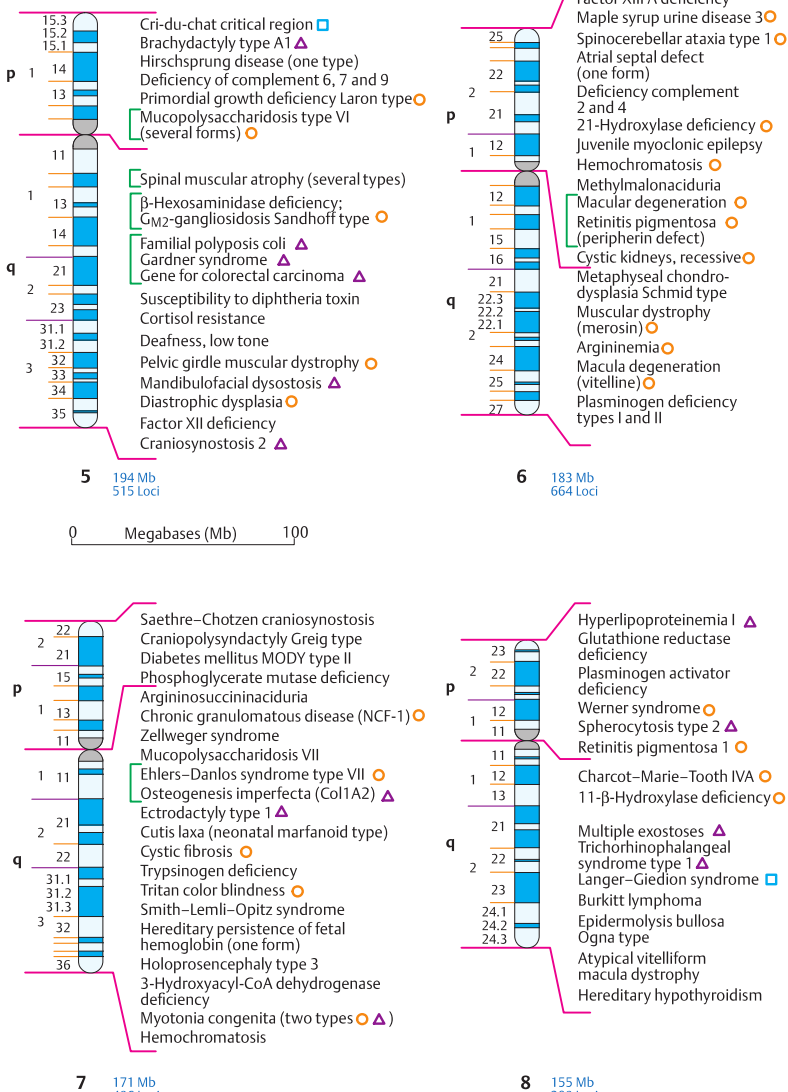
Table. OMIM Statistics: total number of entries: 24,065 (26 May, 2017)

Genes: Autosomal 14,796; X-chromosomal 718; Y-chromosomal 49; mitochondrial 35. Total 15,598	
Phenotypes with molecular basis known: Autosomal 4,649; X-chromosomal 318; Y-chromosomal 4; mitochondrial 31. Total 5,002	
Phenotypes with molecular basis unknown: Autosomal 1,473; X-chromosomal 124; Y-chromosomal 5; mitochondrial 0. Total 1,602	
Phenotypes with suspected mendelian basis: Autosomal 1,672; X-chromosomal 111; Y-chromosomal 2; mitochondrial 68	

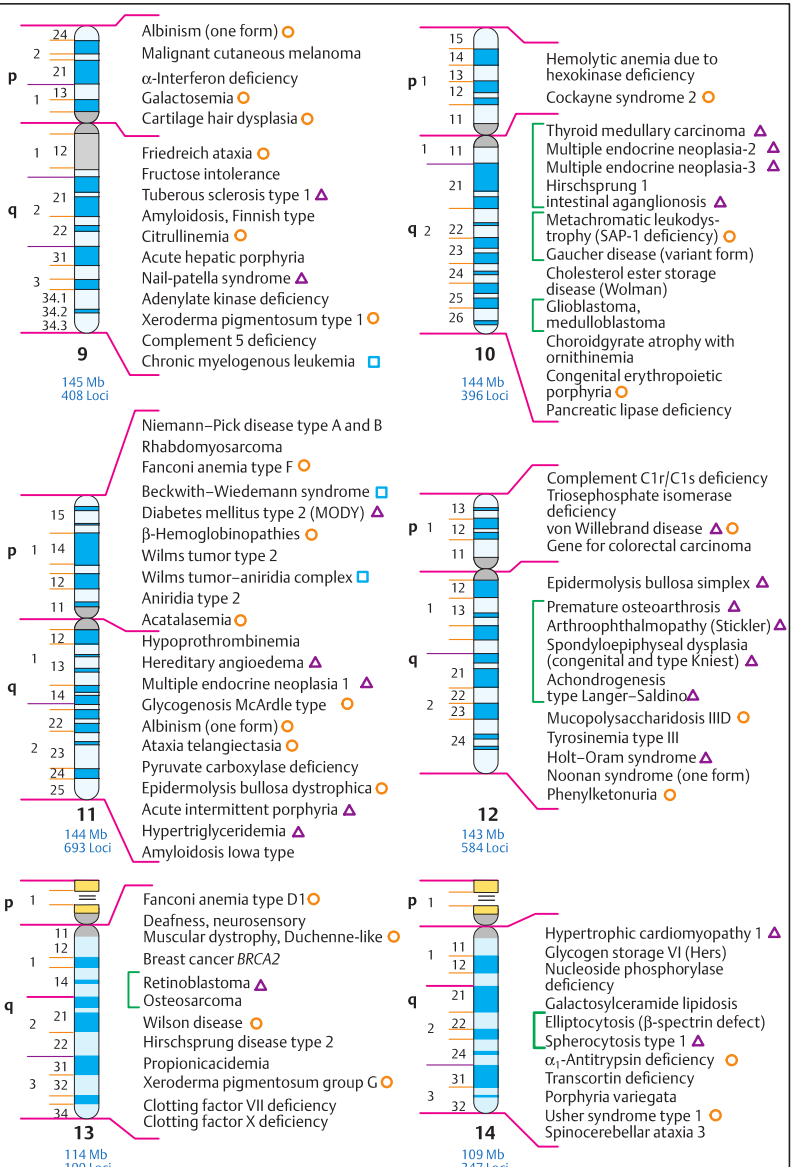
Source: www.ncbi.nlm.nih.gov/omim.



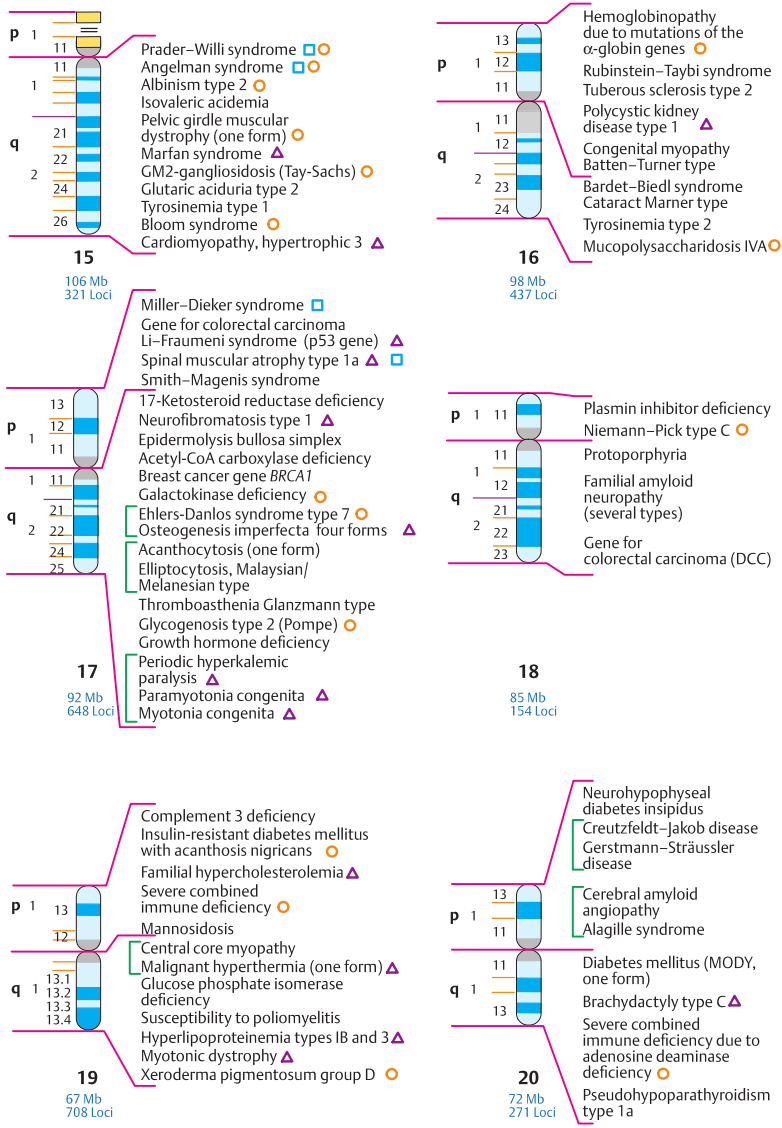
A. Examples of gene loci of genetic diseases in the human genome (chromosomes 1–4)



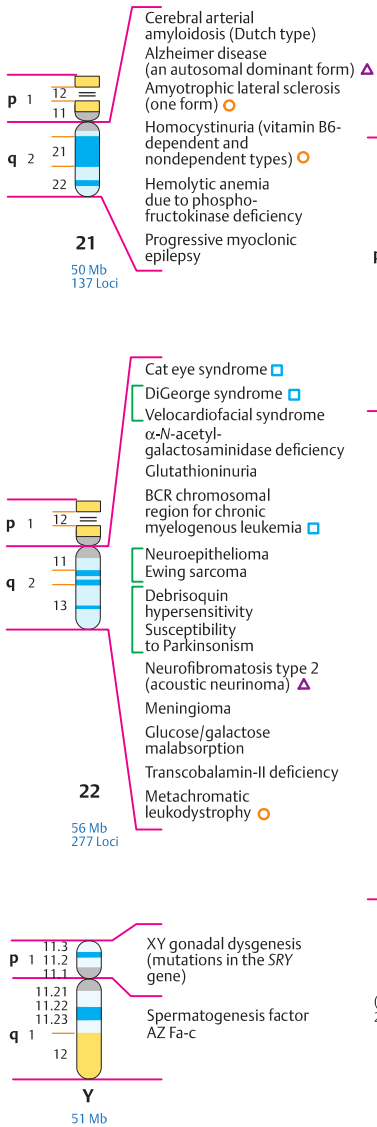
B. Examples of gene loci of genetic diseases in the human genome (chromosomes 5–8)



C. Examples of gene loci of genetic diseases in the human genome (chromosomes 9–14)



D. Examples of gene loci of genetic diseases in the human genome (chromosomes 15–20)



E. Examples of gene loci of genetic diseases in the human genome (chromosomes 21, 22, X, Y)

Chromosomal Locations—Alphabetical List

(Alphabetic list to the maps on pp. 393–397; ch.=chromosome)

Aarskog syndrome (X ch.)
Acanthocytosis (one form) (ch. 17)
Acatalasemia (ch. 11)
Acetyl-CoA carboxylase deficiency (ch. 17)
Achromatopsia (ch. 1)
Achondroplasia (ch. 4)
ACTH deficiency (ch. 2)
Acute hepatic porphyria (ch. 9)
Acute intermittent porphyria (ch. 11)
Acyl-CoA dehydrogenase deficiency (ch. 1)
Adenylate kinase deficiency (ch. 9)
Adrenal cortical hypoplasia with glycerol kinase deficiency (X ch.)
Adrenoleukodystrophy (X ch.)
α-Interferon deficiency (ch. 9)
Agammaglobulinemia (X ch.)
Alagille syndrome (ch. 20)
Albinism (one form) (ch. 9)
Albinism (one form) (ch. 11)
Albinism type 2 (ch. 15)
Albinism–deafness syndrome (X ch.)
Alkaptonuria (ch. 3)
α-N-acetylgalactosaminidase deficiency (ch. 22)
α_1-Antitrypsin deficiency (ch. 14)
α-Fetoprotein deficiency (ch. 4)
Alport syndrome (ch. 2, one of several types on other chromosomes)
Alzheimer disease (one form) (ch. 21)
Amelogenesis imperfecta (X ch.)
Amyloidosis, Finnish type (ch. 9)
Amyloidosis, Iowa type (ch. 11)
Amyotrophic lateral sclerosis (one form) (ch. 21)
Amyotrophic lateral sclerosis, juvenile (ch. 2)
Analbuminemia (ch. 4)
Androgen receptor defect (X ch.)
Angelman syndrome (ch. 15)
Aniridia type 1 (ch. 2)
Aniridia type 2 (ch. 11)
Anterior segmental mesenchymal dysgenesis (ch. 4)
Antithrombin III deficiency (ch. 1)
Apert syndrome (ch. 4)
Apolipoprotein B-100 defect (ch. 2)
Argininemia (ch. 6)
Argininosuccinic aciduria (ch. 7)
Arthro-ophthalmopathy (Stickler syndrome) (ch. 12)
Aspartylglucosaminuria (ch. 4)
Ataxia-telangiectasia (ch. 11)
Atransferrinemia (ch. 3)
Atrial septal defect (one form) (ch. 6)
Atypical vitelliform macular dystrophy (ch. 8)
Bardet–Biedl syndrome (ch. 16, several other types on different chromosomes)
BCR chromosomal region for chronic myelogenous leukemia (ch. 22)
Becker muscular dystrophy (X ch.)
Beckwith–Wiedemann syndrome (ch. 11)
β-Hemoglobinopathies (ch. 11)
β-Hexosaminidase deficiency; GM2 gangliosidosis type Sandhoff (ch. 5)
11-β-Hydroxylase deficiency (ch. 8)
Bloom syndrome (ch. 15)
Brachydactyly type A1 (ch. 5), type C (ch. 20)
Breast cancer gene BRCA1 (ch. 17)
Breast cancer gene BRCA2 (ch. 13)
Burkitt lymphoma (ch. 8)
Carbamoyl phosphate synthetase-I deficiency (ch. 4)
Cardiomyopathy, familial hypertrophic type 3 (ch. 15)
Cartilage–hair dysplasia (ch. 9)
Cat eye syndrome (ch. 22)
C3b-inactivator deficiency (ch. 4)
Central core myopathy (ch. 19)
Cerebral amyloid angiopathy (ch. 20)
Cerebral arterial amyloidosis (Dutch type) (ch. 21)
Cerebrotendinous xanthomatosis (ch. 2)
Charcot–Marie–Tooth neuropathy type1b and type 2 (ch. 17) (spinal muscular atrophy)
Charcot–Marie–Tooth neuropathy type IVa (ch. 8)
Chediak–Higashi syndrome (ch. 1)
Cholesteryl ester storage disease (Wolman) (ch. 10)
Chondrodysplasia punctata (X ch.)
Choroid gyrate atrophy with ornithinemia (ch. 10)
Choroideremia (X ch.)
Chronic granulomatous disease (NCF-1 deficiency) (ch. 7)
Chronic granulomatous disease (NCF-2 deficiency) (ch. 1)
Chronic granulomatous disease (X ch.)
Chronic myelogenous leukemia (ch. 9)
Citrullinemia (ch. 9)
Clotting factor VII deficiency (ch. 13)

Clotting factor X deficiency (ch. 13)
Cockayne syndrome 2 (ch. 10)
Colon cancer, familial nonpolyposis type 1 (ch. 2)
Colorectal adenocarcinoma (ch. 12)
Colorectal carcinoma (ch. 5 and 18)
Colorectal carcinoma/Li–Fraumeni syndrome (ch. 17)
Complement 2 and 4 deficiency (ch. 6)
Complement 3 deficiency (ch. 19)
Complement 5 deficiency (ch. 9)
Complement 6, 7, and 9 deficiency (ch. 5)
Complement 8 deficiency 1 and 2 (ch. 1)
Complement C1r/C1 second deficiency (ch. 12)
Congenital erythropoietic porphyria (ch. 10)
Congenital myopathy, Batten–Turner type (ch. 16)
Cortisol resistance (ch. 5)
Craniopolysyndactyly, Greig type (ch. 7)
Craniosynostosis type 2 (ch. 5)
Creutzfeldt–Jakob disease (ch. 20)
Cri-du-chat critical region (ch. 5)
Crigler–Najjar syndrome (ch. 1)
Crouzon craniofacial dysostosis (ch. 4)
Cutis laxa (neonatal marfanoid type) (ch. 7)
Cystic fibrosis (ch. 7)
Deafness, low-tone (ch. 5)
Deafness, neurosensory (ch. 13)
Debrisoquine hypersensitivity (ch. 22)
Dentinogenesis imperfecta type I (ch. 4)
Diabetes mellitus type MODY (ch. 11, ch 20)
Diabetes mellitus, MODY type II (ch. 7)
Diastrophic dysplasia (ch. 5)
DiGeorge syndrome (ch. 22)
Duchenne muscular dystrophy (X ch.)
Dysfibrinogenemia (ch. 4)
Dyskeratosis congenita (X ch.)
Ectrodactyly type 1 (ch. 7)
Ehlers–Danlos syndrome type 4 (ch. 2)
Ehlers–Danlos syndrome type 7 (ch. 7)
Ehlers–Danlos syndrome type 7 A1 (ch. 17)
Elliptocytosis (β-spectrin defect) (ch. 14)
Elliptocytosis, Malaysian/Melanesian type (ch. 17)
Elliptocytosis-2/recessive spherocytosis (ch. 1)
Epidermolysis bullosa dystrophica (ch. 3)
Epidermolysis bullosa simplex (ch. 12, ch. 17)
Epidermolysis bullosa type Herlitz (ch. 1)
Epidermolysis bullosa type Ogna (ch. 8)
Ewing sarcoma (ch. 22)
Fabry disease (X ch.)
Facioscapulohumeral muscular dystrophy (ch. 4)
Factor V deficiency (ch. 1)
Factor XI deficiency (ch. 4)
Factor XII deficiency (ch. 5)
Factor XIII A deficiency (ch. 6)
Factor XIII B deficiency (ch. 1)
Familial amyloid neuropathy (several types) (ch. 18)
Familial aortic aneurysm (ch. 2)
Familial hypercholesterolemia (ch. 19)
Familial hypertrophic cardiomyopathy (ch. 1)
Familial polyposis coli (ch. 5)
Fanconi anemia type D2 (ch. 6), type F (ch. 15), D1 (ch. 17)
Fragile X syndrome (X ch.)
Friedreich ataxia (ch. 9)
Fructose intolerance (ch. 9)
Fucosidosis (ch. 1)
G6PD deficiency (X ch.)
Galactokinase deficiency (ch. 17)
Galactose epimerase deficiency (ch. 1)
Galactosemia (ch. 9)
Galactosylceramide lipidosis (ch. 14)
Gardner syndrome (ch. 5)
Gaucher disease (ch. 1)
Gaucher disease (variant form) (ch. 10)
Gerstmann–Sträussler disease (ch. 20)
Glioblastoma, medulloblastoma (ch. 10)
Glucose/galactose malabsorption (ch. 22)
Glucose phosphate isomerase deficiency (ch. 19)
Glutaric aciduria type 2 (ch. 15)
Glutathione peroxidase deficiency (ch. 3)
Glutathione reductase deficiency (ch. 8)
Glutathionuria (ch. 22)
Glycogenosis type VII (ch. 1)
Glycogenosis, McArdle type (ch. 11)
Glycogenosis type 2 (Pompe) (ch. 17)
Glycogen storage type 3 (ch. 1)
Glycogen storage type 4 (ch. 3)
Glycogen storage VI (Hers) (ch. 14)
GM1-gangliosidosis (ch. 3)
GM2-gangliosidosis (Tay–Sachs) (ch. 15)
GM2-gangliosidosis, Sandhoff type (ch. 5)
Growth hormone deficiency (ch. 17)
Hemochromatosis (ch. 7)
Hemoglobinopathies caused by mutations of the α-globin genes (ch. 16)
Hemolytic anemia caused by hexokinase deficiency (ch. 10)
Hemolytic anemia caused by phosphofructokinase deficiency (ch. 21)
Hemophilia A (X ch.)
Hemophilia B (X ch.)

Hepatoerythropoietic porphyria (ch. 1)
Hereditary angioedema (ch. 11)
Hereditary congenital hypothyroidism (ch. 8)
Hereditary hypoceruloplasminemia (ch. 3)
Hereditary nephritis (Alport syndrome) (X ch and other loci)
Hereditary nonpolyposis colon carcinoma (ch. 2)
Hereditary persistence of fetal hemoglobin (one form) (ch. 7)
Hirschsprung disease (chs. 10 and 13, and other loci)
Holoprosencephaly type 3 (ch. 7)
Holt–Oram syndrome (ch. 12)
Homocystinuria (B6-responsive and B6-nonresponsive forms) (ch. 21)
Huntington disease (ch. 4)
3-Hydroxyacyl-CoA dehydrogenase deficiency (ch. 7)
21-Hydroxylase deficiency (ch. 6)
Hyperbetalipoproteinemia (ch. 2)
Hyper-IgM immune deficiency (X ch.)
Hyperlipoproteinemia type 1 (ch. 8)
Hyperlipoproteinemia type 1b (ch. 19)
Hyperlipoproteinemia type 3 (ch. 19)
Hypertriglyceridemia (ch. 11)
Hypertrophic cardiomyopathy (ch. 14)
Hypochondroplasia (ch. 4)
Hypohidrotic ectodermal dysplasia (X ch and autosomal types)
Hypophosphatemia (X ch.)
Hypoprothrombinemia (ch. 11)
Immune deficiency, severe combined (ch. 19)
Immunodeficiency due to ADA deficiency (ch. 20)
Impaired hearing (lower frequencies) (ch. 5)
Impaired hearing due to stapes fixation (X ch.)
Infantile ceroid lipofuscinosis, neuronal type (ch. 1)
Infantile hypophosphatasia (ch. 1)
Insulin-resistant diabetes mellitus with acanthosis nigricans (ch. 19)
Interleukin 2 deficiency (ch. 4)
Intestinal aganglionosis (Hirschsprung) (ch. 10 and 13 and other loci)
Isovaleric acidemia (ch. 15)
Juvenile myoclonic epilepsy (ch. 6)
Kallmann syndrome (X ch.)
17-Ketosteroid reductase deficiency (ch. 17)
Lamellar cataract (one type) (ch. 1)
Langer–Giedion syndrome (ch. 8)
Lesch–Nyhan syndrome (X ch.)
Li–Fraumeni syndrome (ch. 17)
Limb–girdle muscular dystrophy 2b (ch. 2)
Lowe syndrome (X ch.)
Lymphoproliferative syndrome (X ch.)
Macular degeneration, vitelline (ch. 6)
Malignant cutaneous melanoma (ch. 9)
Malignant hyperthermia (ch. 19, others)
Mandibulofacial dysostosis (Franceschetti–Klein syndrome) (ch. 5)
Mannosidosis (ch. 19)
Maple syrup urine disease type 2 (ch. 1)
Maple syrup urine disease type 3 (ch. 6)
Marfan syndrome (ch. 15)
Meningioma (ch. 22)
Menkes syndrome (X ch.)
Metachromatic leukodystrophy (ch. 22)
Metachromatic leukodystrophy (SAP-1 deficiency) (ch. 10)
Metaphyseal chondrodysplasia type Schmid (ch. 6)
Methylmalonicaciduria (ch. 6)
Miller–Dieker syndrome (ch. 17)
Morquio syndrome B (ch. 3)
Mucolipidosis types II and III (ch. 4)
Mucopolysaccharidosis type I (Hurler/Scheie) (ch. 4)
Mucopolysaccharidosis type II (X ch.)
Mucopolysaccharidosis type IVa (ch. 16)
Mucopolysaccharidosis type IVb (ch. 3)
Mucopolysaccharidosis type VI (Maroteaux–Lamy) (ch. 5)
Mucopolysaccharidosis type VII (ch. 7)
Multiple endocrine neoplasia type 1 (ch. 11)
Multiple endocrine neoplasia type 2 (ch. 10)
Multiple endocrine neoplasia type 3 (ch. 10)
Multiple exostoses (ch. 8)
Muscular dystrophy, Becker type (X ch.)
Muscular dystrophy, Duchenne type (X ch.)
Muscular dystrophy, Emery–Dreifuss type (X ch.)
Muscular dystrophy, Duchenne-like (ch. 13)
Muscular dystrophy, merosin (ch. 6)
Myotonia congenita (ch. 17)
Myotonia congenita (two types) (ch. 7)
Myotonic dystrophy (ch. 19)
Myotubular myopathy (X ch.)
Nail–patella syndrome (ch. 9)
Nephrogenic diabetes insipidus (X ch.)
Nephronophthisis, juvenile (ch. 2)
Neuroepithelioma (ch. 22)
Neurofibromatosis type 1 (ch. 17)
Neurofibromatosis type 2 (acusticus neurinoma) (ch. 22)

Neurohypophyseal diabetes insipidus (ch. 20)
Niemann–Pick disease type A and type B (ch. 11)
Niemann–Pick type C (ch. 18)
Noonan syndrome (one locus) (ch. 12)
Norrie syndrome (X ch.)
Nucleoside phosphorylase deficiency (ch. 14)
Ocular albinism (X ch.)
Ornithine transcarbamylase deficiency (X ch.)
Oroticacidemia (ch. 3)
Osteogenesis imperfecta (ch. 17 and other loci)
Osteogenesis imperfecta (COL1 A2) (ch. 7)
Osteosarcoma (ch. 13)
Otopalatodigital syndrome type 1 (X ch.)
Pancreatic lipase deficiency (ch. 10)
Paramyotonia congenita (ch. 17)
Pelizaeus–Merzbacher disease (X ch.)
Pelvic girdle muscular dystrophy (ch. 5 and ch. 15)
Periodic hyperkalemic paralysis (ch. 17)
Phenylketonuria (PKU) (ch. 12)
Phosphoglucokinase deficiency (X ch.)
Phosphoglycerate mutase deficiency (ch. 7)
Phosphokinase deficiency (ch. 1)
PKU due to dihydropteridine reductase deficiency (ch. 4)
Plasmin inhibitor deficiency (ch. 18)
Plasminogen activator deficiency (ch. 8)
Plasminogen deficiency, types I and II (ch. 6)
Polycystic kidney disease (ch. 16 and 4, and other loci)
Porphyria cutanea tarda (ch. 1)
Porphyria variegata (ch. 14)
Prader–Willi syndrome (ch. 15)
Primordial growth deficiency, Laron type (ch. 5)
Progressive myoclonic epilepsy (ch. 21)
Propionicacidemia type A (ch. 13)
Propionicacidemia type B (ch. 3)
Protein C deficiency (ch. 2)
Protein S deficiency (ch. 3)
Protoporphyria (ch. 18)
Pseudoaldosteronism (ch. 4)
Pseudohypoparathyroidism type 1a (ch. 20)
Pseudo-Zellweger syndrome (ch. 3)
Red–green blindness (X ch.)
Renal cell carcinoma (ch. 3)
Retinitis pigmentosa (peripherin defect) (ch. 6)
Retinitis pigmentosa type 1 (ch. 8 and other loci)
Retinitis pigmentosa type 2 (X ch.)
Retinitis pigmentosa type 3 (X ch.)
Retinitis pigmentosa type 4 (ch. 3)
Retinoblastoma (ch. 13)
Retinoschisis (X ch.)
Rett syndrome (X ch.)
Rhabdomyosarcoma (ch. 11)
Rieger syndrome (ch. 4)
Rubinstein–Taybi syndrome (ch. 16)
Saethre–Chotzen craniosynostosis (ch. 7)
Sclerotylosis (ch. 4)
Sex reversal (XY females resulting from mutation in the SRY gene) (Y ch.)
Small cell bronchial carcinoma/colorectal cancer (ch. 3)
Smith–Lemli–Opitz syndrome (ch. 7)
Smith–Magenis syndrome (ch. 17)
Spastic paraplegia (X chromosomal form) (X ch.)
Spermatogenesis factors AZFa - c (Y ch.)
Spherocytosis type 1 (ch. 14)
Spherocytosis type 2 (ch. 8)
Spinal muscular atrophy (X ch.)
Spinal muscular atrophy Ia (ch. 17)
Spinal muscular atrophy type IB (ch. 1)
Spinal muscular atrophy IVa (ch. 8)
Spinal muscular atrophy Werdnig–Hoffmann and other types (ch. 5)
Spinocerebellar ataxia type 1 (ch. 6)
Spinocerebellar ataxia type 3 (ch. 14)
Spondyloepiphyseal dysplasia (congenital type) (ch. 12)
Spondyloepiphyseal dysplasia (Kniest type) (ch. 12)
Steroid sulfatase deficiency (ichthyosis) (X ch.)
Sucrose intolerance (ch. 3)
Susceptibility to diphtheria toxin (ch. 5)
Susceptibility to parkinsonism (ch. 22)
Susceptibility to poliomyelitis (ch. 19)
T-cell leukemia/lymphoma (ch. 14)
TFM androgen receptor defect (X ch.)
Thrombasthenia, Glanzmann type (ch. 17)
Thyroid hormone resistance (ch. 3)
Thyroid iodine peroxidase deficiency (ch. 2)
Thyroid medullary carcinoma (ch. 10)
Transcobalamin II deficiency (ch. 22)
Transcortin deficiency (ch. 14)
Trichorhinophalangeal syndrome type 1 (ch. 8)
Triosephosphate isomerase deficiency (ch. 12)
Tritan color blindness (ch. 7)
Trypsinogen deficiency (ch. 7)
Tuberous sclerosis type 1 (ch. 9)
Tuberous sclerosis type 2 (ch. 19)
Tyrosinemia type 1 (ch. 14), type 2 (ch. 16), type 3 (ch. 12)

Usher syndrome type 1 (ch. 14), type 2 (ch. 1, and other loci)
Velocardiofacial syndrome (ch. 22)
Vitelline macular dystrophy (ch. 6)
von Hippel–Lindau syndrome (ch. 3)
von Willebrand disease (ch. 12)
Waardenburg syndrome type 1 (ch. 2 and other loci)
Werner syndrome (ch. 8)
Williams–Beuren syndrome (ch. 4)
Wilms tumor–aniridia complex (ch. 11)
Wilms tumor type 2 (ch. 11)
Wilson disease (ch. 13)
Xeroderma pigmentosum complementation group B (ch. 2)
Xeroderma pigmentosum group C (ch. 3)
Xeroderma pigmentosum group D (ch. 19)
Xeroderma pigmentosum group G (ch. 13)
Xeroderma pigmentosum type 1 (ch. 9)
XY gonadal dysgenesis (Y ch.)
Zellweger syndrome (ch. 7)
Zellweger syndrome type 2 (ch. 1)
Zonular cataract (ch. 1)

Caveat: Numerous disorders with the same or similar phenotype, i.e., clinical manifestation are caused by mutations in different genes at other loci. These may differ in their modes of inheritance. As a result, the genetic risk may differ. This list and the corresponding map positions shown in the plates are not complete, but only examples. For a complete map and list, see OMIM (www.ncbi.nlm.nih.gov/Omim).

Appendix—Supplementary Data

The following tables provide supplementary data for selected plates, and their accompanying text on the topics is listed below.

List of tables

1	The Banding Pattern of Human Chromosomes
2	Nomenclature of Human Chromosome Aberrations
3	Diseases and G Proteins
4	Diseases Related to the Hedgehog Signaling Network
5	Laminopathies
6	Ciliopathies
7	Neurocristopathies
8	Mitochondrial Diseases
9	Inherited Cardiac Arrhythmias
10	Distal Cholesterol Biosynthesis Pathway
11	Mucopolysaccharide Storage Diseases
12	Genetic Immune Deficiency Diseases
13	Oncogenic Chromosome Translocations
14	Genes Involved in Fanconi Anemia
15	Genetic Diseases and FGF Receptors
16	Collagen Molecules and Diseases
17	Disorders of Sexual Development
18	Genes Involved in Inherited Deafness
19	Common Numerical Human Chromosomal Disorders
20	Microdeletion/Microduplication Syndromes

1. The Banding Pattern of Human Chromosomes (p. 146)

Examples of types of commonly used chromosome bands

Type	Goal of application	Principal method
G	Shows euchromatic light bands and heterochromatic dark bands	Pretreatment with trypsin
Q	Darkfield fluorescence; bright fluorescence of heterochromatin, especially Yq	Quinacrine-induced, Hoechst 33258
R	Reverse of G-bands	Pretreatment with alkali at 80–90°C
C	Preferential stain of centromeres and constitutive heterochromatin	Pretreatment with acid and barium hydroxide

Continued ►

Table Continued

Type	Goal of application	Principal method
T	Preferential stain of telomeres	Telomere-specific DNA probes
DAPI	AT-specific fluorescence (4′,6-diamidino-2-phenylindole)	Enhanced chromosome fluorescence
SCN	Analysis of sister chromatid exchanges	5-Bromodeoxyuridine for two divisions
NOR	Selective staining of nucleolus organizing regions of acrocentric chromosomes	Silver nitrate staining
DAPI/ distamycin	Selective staining of pericentric regions of chromosomes 1, 9, 15, 16, Yq	Application of DAPI and distamycin

Source: Schreck RR, Distèche CM. Chromsome banding techniques. Curr Prooc Hum Genet 2001; Chapter 4: Unit 4.2. doi: 10.1002/0471142905.hg0402s00 (online at www.ncbi.nlm.nih.gov/pubmed/18428280). Accessed 6 September 2017

2. Nomenclature of Human Chromosome Aberrations (p. 148)

Human chromosome nomenclature

Abbreviation	Explanation in detail
46,XX	46 chromosomes including two X chromosomes (normal female karyotype)
46,XY	46 chromosomes including one X and one Y chromosome (normal male karyotype)
47,XXY	47 chromosomes including two X chromosomes and one Y chromosome
47,XXX	47 chromosomes including three X chromosomes
47, XY, +21	47 chromosomes including an X and a Y chromosome and an additional chromosome 21 (trisomy 21)
13p	The short arm of a chromosome 13
13q	The long arm of a chromosome 13
13p14	Region 1, band 4 of a chromosome 13
13q14.2	Region 1, band 4.2 of a chromosome 13
2q-	Loss of chromosome material from the long arm of a chromosome 2
t(2;5)(q21;q31)	Reciprocal translocation between a chromosome 2 and a chromosome 5 with breakpoints in 2q21 and 5q31
t(13q14q)	Centric fusion between a chromosome 13 and a chromosome 14
der	A derivative chromosome as a result of a rearrangement
dic(Y)	Dicentric Y chromosome (two centromeres)
del(2)(q21qter)	Deletion in chromosome 2 from q21 until the telomere
fra(X)(q27.3)	Fragile site at q27.3 of an X chromosome
dup(1)	Duplication in a chromosome 1

Continued ►

Table Continued

Abbreviation	Explanation in detail
h	Heterochromatin, constitutive (at centromeres)
i	Isochromosome, e.g., i(Xq) for the long arm of an X chromosome
ins(5;2)(p14; q22q32)	Insertion of the region q22-q32 of a chromosome 2 into region p14 of a chromosome 5
inv(9)(p11q21)	Inversion of a chromosome 9 between p11 and q21 (indicating the breakpoints)
invdup(15)	Inverted duplication in a chromosome 15
mat	Maternal origin
pat	Paternal origin
r(13)	Ring-shaped chromosome 13 (due to a partial deletion)

Source: Data from McGowan-Jordan J, Simons A, Schmid M, eds. ISCN 2016: An International System for Human Cytogenetic Nomenclature. Basel: Karger, 2016.

3. Diseases and G Proteins (p. 188)

Diseases caused by mutations in G-protein-coupled receptors or G proteins[a]

Disease	OMIM	Mutated protein	Type pf inheritance
McCune–Albright syndrome	174800	$G\alpha_s$ subunit of GNAS1 (gain-of-function)	Somatic, 20q13.32
Hypocalciuric hypercalcemia (three genetic types)	145980; 601199 145981; 139313 600740; 602242	Ca^{2+}-sensing receptor; *CASR* gene	AD, 3q13.3-q21.1
Diabetes insipidus	304800 *AVPR2* gene 300538	Arginine vasopressin receptor-2; G-protein subunit $G\alpha_s$ (loss-of-function)	X-chromosomal (Xq28)
Congenital hypothyroidism (six genetic types)	275200	$G\alpha_s$ subunit of TSH receptor (*TSHR* gene) 603372	AR, 14q31.1 (five other autosomal loci)
Congenital hypothyroidism	188545	Thyrotropin-releasing hormone receptor	AD, 8q23.1

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; TSH, thyroid-stimulating hormone.
Note: See OMIM (at: http://www.ncbi.nlm.nih.gov/omim).
[a] Examples only, genetically heterogeneous.

4. Diseases Related to the Hedgehog Signaling Network (p. 192)

Hedgehog and related proteins

Name	Abbreviation	Gene locus	OMIM
Sonic hedgehog	SHH	7q36.3	600725, 142945
Indian hedgehog	IHH	2q35	600726, 112500
Desert hedgehog	DHH	12q13.12	605423, 233420
Hedgehog acetyltransferase	HHAT	1q32.2	605743
Hedgehog interacting protein	HHIP	4q31.21	606178
Smoothened	SMOH	7q32.1	601500, 605462
Patched	PTCH	9q22.32	601309

Source: Data from OMIM (at: http://www.ncbi.nlm.nih.gov/omim).

Mutations and deletions in more than 10 human genes in the hedgehog gene network result in a group of malformation syndromes: holoprosencephaly (OMIM 236100, 142945, and others; see the following table) and accompanying brain malformations. Mutations in PTCH (9q22) cause basal cell nevus syndrome, Gorlin–Goltz type (OMIM 109400). Mutations in SMO (7q31) are found in some basal cell carcinomas and medulloblastomas. The segment polarity gene Ci is an ortholog of the Gli gene family in vertebrates. This family consists of three distinct Gli proteins: Gl1, Gl2, and Gl3. Human GLI1 is expressed in nearly all isolated basal cell carcinomas (OMIM 139150). GLI3 mutations cause cephalopolysyndactyly, Greig type (OMIM 175700, 145400) and Pallister–Hall syndrome (hypothalamic hamartoblastoma; OMIM 146510). Loss-of-function mutations of Ptch and Smo in Drosophila have very similar phenotypes.

Gene loci involved in the causes of holoprosencephaly

Type of holoprosencephaly	Gene locus	OMIM	Gene
HPE1	21q22.3	236100	*HPE1*
HPE2	2p21	157170	*SIX3*
HPE3	7q36	142945	*SHH*
HPE4	18p11.3	142946	*TGIF1*
HPE5	13q32.3	609637	*ZIC2*
HPE6	2q37.1-q37.3	605934	*HPE6*
HPE7	9q22.3	610828	*PTCH1*
HPE8	14q13	609408	*HPE8*
HPE9	2q14.2	610829	*GLI2*

Source: Data from OMIM (at: http://www.ncbi.nlm.nih.gov/omim).

5. Laminopathies (p. 244)

Diseases caused by mutations in genes encoding lamins and lamin-associated proteins

	Gene mutated
Striated muscle diseases	
Emery–Dreifuss muscular dystrophy	*LMNA*, *EDMD*, *SYNE1*, *SYNE2*, *TMEM43*, *TMPO*
Limb-girdle muscular dystrophy	*LMNA*
Dilated cardiomyopathy	*LMNA*, *EDMD*, *SYNE1*, *SYNE2*, *TMEM43*, *TMPO*
Congenital muscular dystrophy	*LMNA*
Heart-hand syndrome	*LMNA*
Lipodystrophy syndromes	
Dunnigan-type familial partial lipodystrophy	*LMNA*
Mandibuloacral dysplasia	*LMNA*, *ZMPSTE24*
Lipoatrophy	*LMNA*
Partial lipodystrophy	*LMNB2*
Accelerated aging disorders	
Atypical Werner syndrome	*LMNA*
Hutchinson–Gilford progeria syndrome	*LMNA*
Restrictive dermopathy	*LMNA*, *ZMPSTE24*
Atypical progeria syndrome	*BANF1*
Peripheral nerve disorders	
Charcot–Marie–Tooth disease	*LMNA*
Adult-onset leukodystrophy	*LMNB1*
Spinocerebellar ataxia type 8	*SYNE1*
Bone diseases	
Buschke–Ollendorff syndrome	*LEMD3*
Melorheostosis	*LEMD3*
Osteopoikilosis	*LEMD3*
Greenberg skeletal dysplasia	*LBR*
Other	
Pelger–Huet anomaly	*LBR*
Arthrogryposis	*SYNE2*

Source: Data from Schreiber KH, Kennedy BK. When lamins go bad: nuclear structure and disease. Cell 2013; 152(6): 1365–1375.

6. Ciliopathies (p. 248)

Examples of genetic disorders with impaired ciliary function

Name	OMIM	Gene	Locus
Alström syndrome	203800 606844	*ALMS1*	2p13.1
Short-rib thoracic dysplasia (17 genetic types)	611263 611177	*IFT80* and others	3q25.33
Bardet–Biedl syndrome (20 genetic types)	209901 209900	*BBS1* and others	11q13.2 and 19 other autosomal loci
Polycystic kidney diseases (3 AD types)	173900 173910 600666	*PKD1* *PKD2* *GANAB*	16p13.3 4q22.1 11q12.3
AR polycystic kidney disease	263200 606702	*PKHD1*	6p12.13-p12.2
Joubert syndrome (27 AR genetic types, one X-chromosomal)	213300	*JBTS1* 27 others	9q34.3 (26 other autosomal loci and Xq22.2
Meckel syndrome (12 genetic AR types)	249000	*MKS1* 609883	17q22 and 9 other autosomal loci
Nephronophthisis (16 genetic AR types)	256100	*NPHP1*	2q13 and 15 other autosomal loci

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.
Source: OMIM (at http://www.ncbi.nlm.nih.gov/omim).

7. Neurocristopathies (p. 250)

Genes involved in Hirschsprung disease

Gene	Location	Main effect	Penetrance (%)	OMIM
RET	10q11.21	Dominant, loss-of-function	50-72	142623 164761
GDNF	5p13.2	Dominant/recessive	Unknown	600837
EDNRB	13q22.3	Recessive	8-85	131244
EDN3	20q13.32	Recessive	Unknown	131242
SOX10	22q13.1	Dominant/recessive	>80	602229
ECE1	1p36.12	Dominant/recessive	Unknown	600423
NTN, *NRTN*	19p13.3	Unknown	Unknown	602018
ZEB2 (formerly *ZFHX1B*, *SIP1*)	2q22.3	Sporadic	Unknown	605802
PHOX2B	4p13	Congenital hypoventilation	Unknown	603851 209880

Note: In addition, susceptibility loci have been identified as noncoding variants in the *RET* gene and at autosomal loci 3p12, 4q31.3; 9q31, and 19q12.

8. Mitochondrial Diseases (p. 260)

Examples of diseases due to mutations or deletions in human mitochondrial DNA[a]

OMIM no.	Disease name	Abbreviation
530000	Kearns–Sayre syndrome (ophthalmoplegia, pigmentary retinal degeneration, cardiomyopathy)	KSS
535000	Leber hereditary optic atrophy	LHON
540000	Mitochondrial myopathy, encephalopathy, lactic acidosis	MELAS
545000	Myoclonus epilepsy with ragged red fibers in muscle	MERRF
551500	Neuropathy, ataxia, retinitis pigmentosa	NARP
603041	Mitochondrial neurogastrointestinal encephalopathy[a]	MNGIE
557000	Pearson marrow-pancreas syndrome	PEAR
515000	Chloramphenicol-induced toxicity	
580000	Deafness, aminoglycoside-induced (mutation A1555G)	
520000	Diabetes–deafness syndrome, maternally transmitted	

Source: Data from OMIM (at: http://www.ncbi.nlm.nih.gov/omim).
[a] This may be autosomal recessive.

9. Inherited Cardiac Arrhythmias (p. 264)

Inherited cardiac arrhythmias (long QT syndromes)[a]

Type	Gene	Locus	Common age of onset	OMIM[b]
LQT 1	*KCNQ1*, *KCNA9*	11p15.5	Childhood (90% by age 20 years)	192500
LQT 2	*KCNH2*	7q35-q36	Young adults (gene formerly known as HERG)	613688
LQT 3	*SCN5A*	3p21-24	Young adults	603830
LQT 4	*ANK2*[c]	4q25	Adults	600919
LQT 5	*KCNE1*	21q22	Children	613695
LQT 6	*KCNE2*	21q22	Adults	613691
LQT 7	*KCNJ2*	17q23	Adults (Andersen syndrome)	170390
LQT 8	*CACNA1*	12p13.33	Syndactyly, immune defect; children and adults	601005
LQT 9	*CAV3*	3p25.3		611818
LQT 10	*SCN4B*	11q23.3		611819
LQT 11	*AKP9*	7q21.2		611820

Continued ►

Table Continued

Type	Gene	Locus	Common age of onset	OMIM[b]
LQT 12	*SNTA1*	20q11.21		612955
LQT 13	*KCNJ5*	11q24.3	Atrial fibrillation	613485
LQT 14	*CALM1*	14q32.11		616247
LQT 15	*CALM2*	2p21		616249
LQT 17	*SCN4B*	11q23.3		611819

Note: Syndromic forms associated with deafness: Romano–Ward syndrome (OMIM 192500), and Jervell and Lange-Nielsen syndrome (OMIM 220400).
[a] Other forms of inherited arrhythmias exist.
[b] OMIM (at: http://www.ncbi.nlm.nih.gov/omim).
[c] Ankyrin-2 (Ank-B in mice).

10. Distal Cholesterol Biosynthesis Pathway (p. 280)

Diseases of the distal (postsqualene) cholesterol biosynthesis pathway

Disease	OMIM	Locus	Gene	Main manifestation
Greenberg dysplasia	215140 600024	1q42.1	*LBR*	Hydrops fetalis, ectopic calcifications, prenatal lethal
CHILD syndrome	308050 300275	Xq28	*NSDHL*	Hemidysplasia, ichthyosiform erythroderma, limb defects
Chondrodysplasia punctata type 2	302960	Xp11.23	*EBP*	Skeletal dysplasias, calcifications over joints, short extremities
Lathosterolosis	607330 602286	11q23.3-q24.1	*SC5DL*	Facial dysmorphism, similar to SLOS, mental retardation
Smith–Lemli– Opitz syndrome (SLOS)	270400 602858	11ql3.4	*DHCR7*	Skeletal anomalies, facial dysmorphia, malformations
Desmosterolosis	602398 606418	1p32.3	*DHCR24*	Facial dysmorphism, short limbs, perinatal lethal

Abbreviations: CYP51, gene encoding cytochrome p51; DHCR7, gene encoding 7-dehydrocholesterol; DHCR24, gene encoding desmosterol reductase; EBP, gene encoding emopamil binding protein; LBR, gene encoding laminin receptor B; NSDHL, gene encoding sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating; SC5DL, gene encoding lathosterol oxidase. (For details, see OMIM and GeneReviews at: http://www.ncbi.nlm.nih.gov/omim.)
Note: For enzymes involved, see page 282.

11. Mucopolysaccharide Storage Diseases (p. 292)

Classification of mucopolysaccharidoses[a]

Type	OMIM	Enzyme defect	Gene locus	Main manifestation
IH Hurler	252800	α-L-Iduronidase	4p16.3	Dysostosis multiplex, developmental delay, corneal clouding
IS Scheie	252800	α-L-Iduronidase	4p16.3	Stiff joints, normal development
II Hunter	309900	Iduronate sulfatase	Xq28	Similar to MPS IH, no corneal clouding
MPS III (Sanfilippo)				
IIIA	252900	Heparin-*N*-sulfatase	17q25.3	Progression mental retardation
IIIB	252920	α-*N*-acetyl-glucosaminidase	17q21.2	Same
IIIC	252930	Acetyl-CoA:α-glucosaminidase	8pll.1	Same
HID	252940	*N*-acetylglucosamine-6-sulfatase	12q14	Same
IV Morquio				
IVA	253000	*N*-acetylgalactosamine-6-sulfatase	16q24.3	Skeletal abnormalities
IVB	253010	β-Galactosidase	3p22.3	Short stature
VI (Maroteaux–Lamy)	253200	*N*-acetylgalactosamine-4-sulfate	5q14.1	Dysostosis multiplex, normal mental development
VII (Sly)	253220	α-*N*-acetyl-glucosaminidase	17q11.21	Dysostosis multiplex, corneal clouding
IX	601492	Hyaluronidase	3p21.31	Soft-tissue masses, short stature

Source: Data adapted from Neufeld EF, Muenzer J. The mucopolysaccharidoses. In: Scriver CR, Beaudet Al, Sly WS, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 3421–3452. (Online at: http://www.ommbid.com/). OMIM: http://www.ncbi.nlm.nih.gov/omim

12. Genetic Immune Deficiency Diseases (p. 310)

Hereditary immune deficiency diseases (examples)

Disease	OMIM	Gene locus	Gene	Inheritance
Disorders of the complement system	106100	11q12.1	*C1NH*	AD
	217000	6p21.33	*C2*	AR
	120700	19p13.3	*C3*	AD
Chronic granulomatous disease	306400	Xp21.1	*CYBB*	X
	300481			
Agammaglobulinemia,[a] Bruton type	300300	Xq22.1	*BTK*	X
Severe combined immunodeficiency (SCID)	300400	Xq13.3	*IL2RG*	X
	601457	11p12	*RAG1/2*	AR
About 20 other types of SCID	300400	Phenotypic Series OMIM		
Adenosine deaminase deficiency	102700	20q13.12	*ADA*	AR
Agammaglobulinemia[b] (μ-chain deficiency)	147020	14q32.33	*IGHM*	AR
DiGeorge syndrome	188400	del22q11.21	*TBX* and others	AD and sporadic
Wiskott–Aldrich syndrome	301000	Xp11.23	*WAS*	AR
Ataxia-telangiectasia	208900	11q22.3	*ATM*	AR

[a] OMIM at: http://www.ncbi.nlm.nih.gov/omim
[b] Agammaglobulinemia types 1–6
(see OMIM 601495 phenotypic series). AD, autosomal dominant; AR, autosomal recessive; X, X-chromosomal.

13. Oncogenic Chromosome Translocations (p. 324)

Examples of oncogenic chromosome translocations leading to fused genes[a]

Translocation	Type of tumor	Genes involved
t(9;22)(q34;q11)	Chronic myeloid leukemia	*ABL/BCR*
t(8;21)(q22;q22)	Acute myeloid leukemia	*AML1/ETO*
t(15;17)(q22;q12)	Promyeloid acute leukemia	*PLM/RARA*
t(14;18)(q32;q21)	Follicular lymphoma	*BCL2/IgH*
t(14;19)(q32;q13)	B-cell lymphocytic leukemia	*BCL3/IgH*
t(8;14)(q24;q32)	Burkitt lymphoma, B-cell ALL	*MYC/IgH*
t(11;14)(q13;q32)	Mantle cell lymphoma	*BCL/IgH*
t(1;7)(p34;q35)	T-cell acute lymphocytic leukemia	*LCK/TCRB*
t(4;11)(q21;q23)	Acute lymphocytic leukemia	*MLL/ALL1/HRX*
t(3;21)(q26;q22)	Acute myeloid leukemia	*AML1/EAP/EV11*
t(1;14)(p22;q32)	Mucosa-associated lymphoma (MALT)	*BCL10*
t(21;22)(q22;q12)	Ewing sarcoma	*EWS/ERG*
t(11;22)(q24;q12)	Ewing sarcoma	*EWS/FLI1*

Source: Data from Morin PJ, Trent JM, Collins FS, Vogelstein B. Cancer genetics. In: Longo DL, Fauci AS, Kasper DL, Hauser SL, jameson JL, Loscalzo J, eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill, 2012: 663–672, and OMIM (http://www.ncbi.nlm.nih.gov/omim).

14. Genes Involved in Fanconi Anemia (p. 330)

Gene	Locus	OMIM[a]
FA-A	16q24.3	227650
FA-B	Xp22.2	300514
FA-C	9q22.32	227645
FA-D1[b]	13q13.1	605724
FA-D2	3p25.3	227646
FA-E	6p21.31	600901
FA-F	11pl4.3	603467
FA-G	9p13.3	614082
FA-I	15q26.1	609053
FA-J	17q23.2	609054
FA-L	2p16.1	614083

Continued ►

Table Continued

Gene	Locus	OMIM[a]
FA-M	14q21.2	609644
FA-N[c]	16p12.2	610832
FA-O[d]	17q22	613390
FA-P	16p13.3	613951
FA-Q	16p13.12	615272
FA-R	15q15.1	617244
FA-T	1p32.1	616435

Source: Data from Schindler D, Hoehn H, eds. Fanconi Anemia. Basel: Karger, 2007.
[a] OMIM at: http://www.ncbi.nlm.nih.gov/omim.
[b] Same as *BRCA2*.
[c] Same as *PALB2*.
[d] Same as *RADC51*.

15. Genetic Diseases and FGF Receptors (p. 340)

Examples of genetic disorders due to mutations in FGF receptors[a]

Gene	Location	Disease	Main manifestation	OMIM[b]
FGFR1	8p11.23	Pfeiffer syndrome	Craniosynostosis, broad thumbs	101600
FGFR2	10p26.13	Apert syndrome	Craniosynostosis, fused digits	101200
		Crouzon syndrome	Craniosynostosis, ocular proptosis	123500
		Pfeiffer syndrome	Craniosynostosis, broad thumbs	101600
		Saethre–Chotzen syndrome[c]	Craniostenosis, brachycephaly, hand and feet anomalies	101400
		Antley–Bixler syndrome	Craniostenosis and radiohumeral synostosis	207410 176943
FGFR3	4p16.3	Achondroplasia	Short stature, bone dysplasias	100800
		Hypochondroplasia	Mild form of achondroplasia	146000
		Muenke syndrome	Asymmetric coronal stenosis	602849

[a] Many more exist; some with unique mutations.
[b] OMIM at: http://www.ncbi.nlm.nih.gov/omim.
[c] Other forms exist (OMIM 101400 due mutations in *TWIST* 601622).

16. Collagen Molecules and Diseases (p. 344)

Important types of human collagen and diseases caused by mutations in their genes

Type	Molecular structure	Gene	Gene locus	Disease	OMIM no.[a]
I	[α1(I)$_2$ α 2(II)]	*COL1A1*	17q21.33	Osteogenesis imperfecta type I–IV, Ehlers–Danlos syndrome	120150
		COL1A2	7q21.3	Ehlers–Danlos syndrome, OI II–IV	120160 130000
II	[α1(II)3]	*COL2A1*	12q13.11	Stickler syndrome Spondyloepiphyseal dysplasia Achondrogenesis Others	120140 108300 183900 200600
III	[α1(III)$_3$]	*COL3A1*	2q32.2	Ehlers–Danlos syndrome IV	130050
IV	[α1(IV)α2(IV)] and others	*COL4A1*, *A2*	13q34	Alport syndrome autosomal	203780
		A3,*A4*	2q36.3	Alport syndrome	301050
		A5,*A6*	Xq22	X-chromosomal	
V	[α 1(V)$_2$ α 2(V)]	*COL5A1* *COL5A2*	9q34.3 2q31	Ehlers–Danlos syndrome I and II	130000

Source: Data from Byers PH. Disorders of collagen synthesis and structure. In: Scriver CR, Beaudet Al, Sly WS, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 5241– 5285. Online at: http://www.ommbid.com.
[a] OMIM (at http://www.ncbi.nlm.nih.gov/omim).

17. Disorders of Sexual Development (p. 366)

Overview of disorders of sex differentiation

1	Defects of sex determination caused by mutation or structural aberration of the *SRY* gene at Yp11.2 (480000): XX sex reversal 1(400045), 2 (278850), 3 (300833), XY sex reversal (400044, gonadal dysgenesis, 10 genetic types)
2	Disorders of testis development (*SF1*, *DAX1*, *WNT4*, *SOX9*, and other genes)
3	Defects of androgen biosynthesis (e.g., 21-hydroxylase deficiency, 201910)
4	Defect in steroid 5α-reductase (dihydrotestosterone deficiency, 264600, 607306)
5	Defects of androgen receptor ("testicular feminization," 300068)
6	Defects of the müllerian inhibition (hernia uteri syndrome, 261550, two types)
7	XO/XY gonadal dysgenesis
8	Turner syndrome (45,X), Klinefelter syndrome (47,XXY)
9	True hermaphroditism XX/XY

18. Genes Involved in Inherited Deafness (p. 376)

Examples of genes and proteins involved in hereditary deafness

Type of protein	Main function	Gene	DFN type	OMIM	Mouse mutant
Cytoskeletal proteins					
Myosin 6	Motor protein	*MYO6*	DFNB37/A22	600970	Snell waltzer
Myosin 7A	Motor protein	*MYO7A*	DFNB2/A11	276903	Shaker-1
Myosin 15	Motor protein	*MYO15*	DFNB3	600316	Shaker-2
Ion transporters					
Connexin 26	Gap junction	*GJB2/B3 CX26*	DFNB1/A3	220290	
Connexin 30	Gap junction	*BJB6/CX30*	DFNB1/A3	604418	
KCNQ4	K^+ channel	*KCNQ4*	DFNA2	600101	
Pendrin	Iodide–chloride	*SLC26A4*	DFNB4	605646 274600	
Structural proteins					
α-Tectorin	Tectorial membrane	*TECTA*	DFNB21/A8/ A12	602574	
Collagen XI	Extracellular matrix	*COL11 A2*	DFNB53	609706	
Cochlin	Extracellular matrix	*COCH*	DFNA9	603196	
POU3F4		*POU3F4*	DFN3	300039	
Mitochondrial					
12S RNA	Aminoglycoside-induced deafness	*MTRNR1 MTCO1*	DFNA5	580000 610230	
Unknown					
Diaphanous	Actin polymerization in hair cells	*DIAPH1*	DFNA1	124900 602121	

Source: Data from Petit C, Levilliers J, Marlin S, Hardelin JP. Hereditary hearing loss. In: Scriver CR, Beaudet Al, Sly WS, et al., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York, NY: McGraw-Hill, 2001: 6281–6328 (online at: http://www.ommbid.com); and Petersen MB, Willems PJ. Non-syndromic, autosomal-recessive deafness. Clin Genet 2006; 69: 371–392 (OMIM at: http://www.ncbi.nlm.nih.gov/omim).

19. Common Numerical Human Chromosomal Disorders (p. 382)

Examples of common numerical human chromosomal disorders[a]

Aberration	Frequency (liveborn)	Main manifestations
Trisomy 21	1:650	Muscular hypotonia, flat face, oblique eye lids, congenital heart defect (60%), other congenital malformations, developmental delay
Trisomy 18	1:5,000	Characteristic face, muscular hypertonia, congenital malformations (heart 90%), severe developmental impairment, lethality during first year 90%
Trisomy 13	1:8,000	Arrhinencephaly, anophthalmia, microphthalmia, cleft lip/palate, heart defect, severe developmental impairment, lethality during first year 90%
Monosomy X females (mosaic 45X/46,XX common) Turner syndrome	1:2,500 females	Short stature, congenital lymphedema and/or webbed neck (variable), coarctation aortae (45%), streak ovaries, other malformations; other manifestations, highly variable
47,XXY Klinefelter syndrome	1:600 males	Postpubertal delay of sexual maturation, small testes, hypogonadism, infertility (highly variable)
47,XXX	1:1,000 females	No recognizable phenotype; variable, usually mild neuromotor deficits
47,XYY	1:1,000 males	No recognizable phenotype; variable, mild developmental delay

[a] Several other rare numerical aberrations occur with variable manifestations and developmental impairment: XXXX, XXXXX, XXYY, XXXY, XXXXY (see references on p. 384).

20. Microdeletion/Microduplication Syndromes (p. 386)

Examples of microdeletion/microduplication syndromes[a]

Chromosomal region	Name	Comments	OMIM
1p36 3q29	1p36 Deletion syndrome Microdeletion 3q29 syndrome	See p. 386	607872 609425
4p16.3	4p- or Wolf–Hirschhorn syndrome	See p. 386	194190
5p15.2-15.3	5p- or cri-du-chat syndrome	Familial occurrence by translocation in 12–15%	123450
5q35	Sotos syndrome	Overgrowth, retardation, seizures; NSD1 gene deletions; frequent in Japan	117550
7q11.23	Williams–Beuren syndrome	Elastin gene and another gene involved; deletion shown in 70%	194050
8q24.12	Langer–Giedion syndrome (I and II)	Sparse hair, bulbous nose, mental retardation	190350
11p13	Wilms tumor–aniridia–genitourinary anomalies (WAGR)	WT1 and PAX6 genes involved	194072
11p15.5 dup	Beckwith–Wiedemann syndrome	Macrosomia, macroglossia	130650
15q11-13	Prader–Willi syndrome	Paternal chromosome 15 involved (see p. 368)	176270
15q11-13 15q13.3	Angelman syndrome Microdeleton 15q13.1 syndrome	Maternal chromosome 15 involved, UBE3A gene mutations (see p. 368)	105830 612001
16p11.2 16p13.3	Microdeletion syndrome 16p11.2 Rubinstein–Taybi syndrome	Gene encoding CREB binding protein involved	611913 180849 600140
16pter-p13.3	ATR-16 syndrome	Mental retardation, deletion hemoglobin-α genes	141750
17p13.3	Miller–Dieker syndrome	Lissencephaly, L1S1 gene deletion in about 90%	247200

Continued ►

Table Continued

Chromosomal region	Name	Comments	OMIM
17p11.2 17q11.2	Smith–Magenis syndrome NF1 microdeletion	Complex malformation syndrome NF1	182290 613675
20p12.1	Alagille syndrome	Arteriohepatic dysplasia and other systemic manifestations, *JAG1* gene mutations	118450
22q11	DiGeorge/Shprintzen syndrome	Immune defects, neonatal hypocalcemia, congenital heart defects, wide clinical spectrum, *TBX1* gene deletion in 70–90%	192430
Xq21.1	ATRX syndrome	Mental retardation, α-thalassemia, dysmorphias	301040

[a] Other microdeletions have been observed at 1q21.1, 1q41-42, 2p15-16.1, 2q23.1, 5q14.3, 12q14, 15q13.3, 15q24.1, 15q26, 16p11.2, 16p13.11, 17q21.3, 19p13.11, 16q11.2, 17q12, 17q21.31, 18q21, and others. (See Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. Ann Rev Genet 2011; 45: 203–226. Also see: DECIPHER GRCh37. List of 70 microdeletion syndromes (at: https://decipher.sanger.ac.uk/syndromes). Accessed July 5, 2017. OMIM (at: http://www.ncbi.nlm.nih.gov/omim).

Definition of Terms in Genetics and Genomics—Glossary

Acentric Refers to a chromosome or chromatid without a centromere.

Acrocentric (White 1945) Refers to a chromosome with a centromere that lies very close to one of the ends, dividing the chromosome into a long arm and a very short arm.

Actin A structural protein interacting with many other proteins. In muscle cells, as F actin, it interacts with myosin during contraction.

Allele (Johannsen 1909) **or allelomorph** (Bateson and Saunders 1902) One of several alternative forms of a gene at a given gene locus.

Allelic disorders A group of disorders resulting from different mutations in the same gene.

Allelic exclusion Expression from one allele only.

Allozygous Refers to a gene locus with alleles of independent origin (see autozygous).

Alpha helix A folding pattern of proteins. A linear sequence of amino acids folds into a helix (compare with beta sheet).

Alternative splicing Production of different mRNAs from one transcript.

Alu sequences A family of related DNA sequences that are each ~300 base pairs long and containing the recognition site for the Alu restriction enzyme; ~1.2 million copies of Alu sequences are dispersed throughout the human genome. A full-length Alu repeat is a dimer of about 280 bp: 120 bp for each monomer, followed by a short sequence rich in A residues. They are asymmetric: the repeat to the right contains an internal 32-bp sequence, while the other does not.

Amber codon The stop codon UAG (a word play on the discoverer Bernstein [amber in German]).

Ames' test A mutagenicity test performed with a mixture of rat liver and mutant bacteria.

Amino acid An organic compound with an amino ($-NH_2$) and a carboxyl ($-COOH$) group.

Aminoacyl tRNA A transfer RNA carrying an amino acid.

Amplification Production of additional copies of DNA sequences.

Amplicon Fragment of DNA amplified by a specific PCR (polymerase chain reaction) reaction.

Anaphase (Strasburger 1884) A stage of mitosis and of meiosis I and II. Characterized by the movement of homologous chromosomes (or sister chromatids) toward opposite poles of the cell division spindle.

Ancestral origin Refers to genetic material of common origin from previous generations.

Aneuploidy (Täckholm 1922) Deviation from the normal number of chromosomes by gain or loss (see trisomy and monosomy).

Aneusomy Deviation from the normal presence of homologous chromosomal segments. Aneusomy by recombination refers to the duplication/deficiency resulting from crossing-over within an inversion (inverted region).

Anneal Hybridizing complementary single strands of nucleic acid to form double-stranded molecules (DNA with DNA, RNA with RNA, or DNA with RNA).

Antibody A protein (immunoglobulin) that recognizes and binds to an antigen as part of the immune response.

Anticipation A tendency of a disease to increase in severity in successive generations.

Anticodon A trinucleotide sequence in tRNA that is complementary to a codon for a specific amino acid in mRNA.

Antigen A molecule that induces an immune response.

Antiparallel Double-stranded DNA or RNA running in opposite directions.

Antisense RNA An RNA strand that is complementary to normal mRNA. Natural antisense RNAs made from the nontemplate strand of a gene may regulate gene expression by preventing it from being used as a template for normal translation.

Apoptosis Programmed cell death.

Archaea Archaebacteria, one of the three evolutionary lineages of organisms living today, a domain distinct from prokaryotes and eukaryotes.

ARMS Amplification refractory mutation system. A genetic test using allele-specific PCR.

Array CGH See CGH. A genetic test based on a comparative analysis by hybridization of a large number of DNA or RNA samples of different origin (test sample and reference sample) on a small surface (array) in the search for deletions or duplications in the genome (see copy number variation).

Ascertainment The mode of identification of individuals for a genetic study.

Ashkenazi The Jewish population descended from Eastern Europe and medieval Jewish communities that lived along the river Rhine and in the Alsace region (Ashkenaz is the medieval Hebrew name for this region). (Compare with Sephardim.)

ASO Allele-specific oligonucleotides.

Association Different alleles occurring together more often than expected from their individual frequencies (as opposed to linkage).

Assortative mating Nonrandom selection of mates based on shared phenotypic features, as opposed to panmixia.

Attenuator A DNA sequence regulating the termination of transcription, involved in controlling the expression of some operons in bacteria.

Aurignacian Tools and artwork associated with upper paleolithic fossils of ancestral modern humans in Europe. Named after Aurignac in the Pyrenees, where it was first discovered. (Compare with Mousterian.)

Australopithecus (Dart 1924) Ancestral forms of early hominids in Africa. Derived from australis (southern) and pithecus (ape).

Autonomously replicating sequence (ARS) A DNA sequence that is required to induce replication.

Autoradiography (Lacassagne and Lattes 1924) Photographic detection of a radioactive substance incorporated into cells or tissue. The distribution of the radioactively labeled substance can be demonstrated, e.g., in tissue, cells, or metaphase chromosomes, by placing a photographic film or photographic emulsion in close contact with the preparation.

Autosome (Montgomery 1906) Any chromosome except a sex chromosome (the latter is usually designated X or Y). Autosomal refers to genes and chromosomal segments that are located on autosomes.

Auxotrophy (Ryan and Lederberg 1946) Refers to cells or cell lines that cannot grow on minimal medium unless a certain nutritive substance is added (compare with Prototrophy).

Backcross Cross of a heterozygous animal with one of its homozygous parents. In a double backcross, two heterozygous gene loci are involved.

BAC Bacterial artificial chromosome; a synthetic DNA molecule that contains bacterial DNA sequences for replication and segregation (see YAC).

Bcl-2 family A family of proteins localized to mitochondria involved in regulation of apoptosis.

Bacteriophage A virus that infects bacteria. Usually abbreviated to phage.

Banding pattern (Painter 1939) A specific staining pattern of a chromosome consisting of alternating light and dark transverse bands. Each chromosomal segment of homologous chromosomes shows the same banding pattern, characterized by the distribution and size of the bands, which can be used to identify that segment. The term was introduced in 1939 for the linear pattern of bands in polytene chromosomes of certain Diptera (mosquitoes, flies). Each band is defined relative to its neighboring bands. The sections between bands are interbands.

Barr body See X chromatin.

Base pair (bp) In double-stranded DNA, bp refers to two nucleotide bases—one a purine, the other a pyrimidine—lying opposite each other and joined by hydrogen bonds. Normal base pairs are AT and CG. Other pairs can form in ribosomal RNA.

B cells B lymphocytes.

Beta sheet A structural motif of proteins. Different sections of the polypeptide run alongside of each other.

Bimodal distribution A frequency distribution curve with two peaks. It may indicate two different phenotypes distinguished on a quantitative basis.

Bivalent (Haecker 1892) Pairing configuration of two homologous chromosomes during meiosis in the pachytene stage. As a rule, the number of bivalents corresponds to half the normal number of chromosomes in diploid somatic cells. Bivalents are the prerequisite for recombination by crossing-over of nonsister chromatids. A trisomic cell forms a trivalent of the three chromosomes.

Breakage–fusion–bridge cycle Refers to a broken chromatid that fuses to its sister, forming a bridge.

Breakpoint Site of a break in a chromosomal alteration, e.g., translocation, inversion, or deletion.

Cadherins Dimeric cell adhesion molecules.

Carcinogen A chemical substance that can induces cancer.

Caspase A member of the family of specialized cysteine-containing aspartate proteases involved in apoptosis (programmed cell death).

CT (or CAAT) box A regulatory DNA sequence in the 5′ region of eukaryotic genes; transcription factors bind to this sequence.

Catenate A link between molecules.

C-bands Specific staining of the centromeres of metaphase chromosomes.

cDNA Complementary DNA synthesized by the enzyme reverse transcriptase from RNA as the template.

CD region Common docking, a region involved in binding to a target protein.

Cell cycle (Howard and Pelc 1953) Life cycle of an individual cell. In dividing cells, the following four phases can be distinguished: G_1 (interphase), S (DNA synthesis), G_2, and mitosis (M). Cells that do not divide are in G_0 phase.

Cell hybrid A somatic cell generated by fusion of two cells in a cell culture. It contains the complete or incomplete chromosome complements of the parental cells.

Cell organelle Defined structural and functional unit within a cell, e.g., mitochondrion, ribosome, endoplasmic reticulum, Golgi's apparatus, lysosome.

CEPH A set of three-generation families with known DNA marker genotypes (Centre d'Étude du Polymorphisme Humain in Paris), introduced by J. Dausset in 1986.

Centimorgan A unit of length on a linkage map (100 centimorgans, cM = 100 Morgan). The distance between two gene loci in centimorgans corresponds to their recombination frequency expressed as percentage, i.e., one centimorgan corresponds to 1% recombination frequency. Named after Thomas H. Morgan (1866–1945), who initiated the classic genetic experiments on *Drosophila* in 1910.

Centriole Small cylinder of microtubules.

Centromere (Waldeyer 1903) Chromosomal region to which the spindle fibers attach during mitosis or meiosis. It appears as a constriction at metaphase. It contains chromosome-specific repetitive DNA sequences.

CGH Comparative genomic hybridization. Compares whole genomes to identify regions with a loss or additional DNA.

Chaperone A protein needed to assemble or fold another protein correctly.

Chiasma (Janssens 1909) Cytologically recognizable region of crossing-over in a bivalent. In some organisms, the chiasmata move toward the end of the chromosomes (terminalization of the chiasmata) during late diplotene and diakinesis (see meiosis). The average number of chiasmata in autosomal bivalents is ~52 in human males, and ~25–30 in females. The number of chiasmata in man was first determined in 1956 in the paper that confirmed the normal number of chromosomes in man by Ford and Hamerton.

Chimera (Winkler 1907) An individual or tissue that consists of cells of different genotypes of prezygotic origin.

ChIP Chromatin immunoprecipitation. A method to detect DNA sequences that bind to a specific protein in chromatin.

Chromatid (McClung 1900) Longitudinal subunit of a chromosome resulting from chromosome replication; two chromatids are held together by the centromere and are visible during early prophase and metaphase of mitosis, and between diplotene and the second metaphase of meiosis. Sister chromatids arise from the same chromosome; nonsister chromatids are the chromatids of homologous chromosomes. After division of the centromere in anaphase, the sister chromatids are referred to as daughter chromosomes. A chromatid break or a chromosomal aberration of the chromatid type affects only one of the two sister chromatids. It arises after the DNA replication cycle in the S phase (see Cell cycle). A break that occurs before the S phase affects both chromatids and is called an isolocus aberration (isochromatid break).

Chromatin (Flemming 1882) The stained material observed in interphase nuclei. It is a general term for packaged DNA, composed of DNA, basic chromosomal proteins (histones), nonhistone chromosomal proteins, and small amounts of RNA.

Chromatin remodeling The energy-dependent displacement or reorganization of nucleosomes for transcription or replication.

Chromosome (Waldeyer 1888) The gene-carrying structures that are composed of chromatin and are visible during nuclear division as threadlike or rodlike bodies ("stained thread" from Greek).

Chromosome painting A staining method to distinguish individual chromosomes.

Chromosome walking Sequential isolation of overlapping DNA sequences to find a gene on the chromosome studied.

Cis-acting Refers to a regulatory DNA sequence located on the same chromosome (cis), as opposed to trans-acting over a distance from other locations.

Cis/trans (Haldane 1941) In analogy to chemical isomerism, refers to the position of genes of double heterozygotes (heterozygotes at two neighboring gene loci) on homologous chromosomes. When two certain alleles, e.g., mutants, lie next to each other on the same chromosome, they are in cis position. If they lie opposite each other on different homologous chromosomes, they are in a trans position. The cis/trans test (Lewis, 1951; Benzer, 1955) uses genetic methods (genetic complementation) to determine whether two mutant genes are in the cis or in the trans position. With reference to genetic linkage, the expressions cis and trans are analogous to the terms coupling and repulsion (q.v.).

Cistron (Benzer 1955) A functional unit of gene effect as represented by the cis/trans test. If the phenotype is mutant with alleles in the cis position and the alleles do not complement each other (genetic complementation), they are considered alleles of the same cistron. If they complement each other, they are considered to be nonallelic. This definition by Benzer was later expanded by Fincham in 1959: accordingly, a cistron now refers to a segment of DNA that encodes a unit of gene product. Within a cistron, mutations in the trans position do not complement each other. Functionally, the term cistron can be equated with the term gene.

Clade A group of organisms evolved from a common ancestor.

Clathrin (Pearse 1975) A protein interacting with adaptor proteins to form the coated vesicles that bud from the cytoplasm during endocytosis.

Clone (Webber 1903) A population of molecules, cells, or organisms that have originated from a single cell or a single ancestor and are identical to it and to each other.

Clonal selection and expansion Specific increase in number of a single cell, e.g., lymphocytes with a specific antigen receptor or cancer cells.

Cloning efficiency A measure of the efficiency of cloning individual mammalian cells in culture.

Cloning vector A plasmid, phage, or bacterial or yeast artificial chromosome (BAC, YAC) used to carry a foreign DNA fragment for the purpose of cloning (producing multiple copies of the fragment).

Coalescence time The time going back to the most recent common ancestor (see Divergence time).

Coding strand of DNA The strand of DNA bearing the same sequence as the RNA strand (mRNA) that is used as a template for translation (sense RNA). The other strand of DNA, which directs synthesis of the mRNA, is the template strand (see antisense RNA).

Coding RNA RNA involved in transcription (see noncoding RNA).

Codominant Expression of two dominant traits together, e.g., the AB blood group phenotype (see Dominant).

Codon (Brenner and Crick 1963) A sequence of three nucleotides (a triplet) in DNA or RNA that codes for a certain amino acid or for the terminalization signal of an amino acid sequence.

Coefficient of inbreeding An expression for the proportion of loci that are homozygous by descent from an ancestor, observed by parental consanguinity.

Coefficient of relationship The proportion of loci that is homozygous for the same allele by descent from a common ancestor.

Co-factor A molecule or metal required for a biological activity of a protein.

Cohesin A protein complex of four subunits (Scc1, Scc3, Smc1, Smc3) regulating the separation of sister chromatids during cell division.

Coiled-coil A stable rodlike structural motif in proteins formed between two or more α-helices.

Collinear The 1:1 representation of triplet nucleotides in DNA and the corresponding sequence of amino acids.

Common disease—common variant hypothesis Genetic influences on common diseases may be attributed to common variants present in 1–5% of the population.

Comparative genomic hybridization (CGH) A test comparing the genome of an individual with a control to assess differences in the amount of DNA that result from a duplication or a deletion in one or several chromosomes (see Array CGH).

Complementary Two nucleic acids with identical sequences in opposite directions. This enables them to anneal and form a double-stranded molecule.

Complementation, genetic (Fincham 1966) Complementary effect of (restoration of normal function by) double mutants at different gene loci (see complementation groups in xeroderma pigmentosum, or Fanconi's anemia).

Compound heterozygote Two different mutant alleles at the same gene locus.

Concatemer An association of DNA molecules with complementary ends linked head to tail and repeated in tandem. Formed during replication of some viral and phage genomes.

Concordance The occurrence of a trait or a disease in both members of a pair of twins (monozygotic or dizygotic) (see Discordance).

Condensation Refers to the density of chromatin in a chromosome.

Condensin Protein complexes (condensin I and condensin II), involved in the chromosome assembly for cell division (chromosome condensation).

Conformation The three-dimensional structure of a complex molecule required for biological function.

Conjugation (Lederberg, Hayes, Cavalli 1953) The transfer of DNA from one bacterium to another.

Consanguinity Blood relationship. Two or more individuals are referred to as **consanguineous** (related by blood) if they have one or more ancestors in common. A quantitative expression of consanguinity is the coefficient of inbreeding (q.v.).

Consensus sequence A corresponding or identical DNA sequence in different genes or organisms.

Conserved in evolution Refers to genes or parts of chromosomes that have been maintained in the course of evolution because of their functional importance.

Contig A series of overlapping DNA fragments (contiguous sequences).

Contiguous gene syndrome A disease resulting from pathological changes in adjacent genes with different functions.

Copy number variation (CNV) Differences in the number of a defined DNA sequence repeats in the genome of different individuals.

Cosmid A plasmid carrying the cos site (q.v.) of a phage in addition to sequences required for division. Serves as a cloning vector for DNA fragments up to 40 kb.

Cos site A restriction site required of a small strand of DNA to be cleaved and packaged into the λ phage head.

Coupling (Bateson et al. 1905) A cis configuration (q.v.) of double heterozygotes.

Covalent bond A stable chemical bond holding molecules together by sharing one or more pairs of electrons (as opposed to a noncovalent hydrogen bond).

CpG island A stretch of 1–2 kb in mammalian genomes that is rich in unmethylated CpG doublets; usually at the 5′ end of a gene.

CRISPR **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats. Short DNA sequences in bacteria and archaea using small noncoding RNAs (ncRNA) to recognize and destroy invading viral genomes by complementary base pairing and targeted nuclease digestion.

Crossing-over (Morgan and Cattell 1912) The exchange of genetic information between two homologous chromosomes by chiasma formation (q.v.) in the diplotene stage of meiosis I; leads to genetic recombination of neighboring (linked) gene loci. Unequal crossing-over (Sturtevant 1925) results from mispairing of the homologous DNA segments at the recombination site. It results in structurally altered DNA segments or chromosomes, resulting in a duplication in one and a deletion in the other. Crossing-over may also occur in somatic cells (Stern 1936).

Cyclic AMP (cAMP) Cyclic adenine monophosphate, a second messenger produced in response to stimulation of G-protein-coupled receptors.

Cyclin A protein involved in cell cycle regulation.

Cytokine A small secreted molecule that can bind to cell surface receptors on certain cells to trigger their proliferation or differentiation.

Cytoplasmic inheritance Transmission of genetic information located in mitochondria. Since sperm cells do not contain mitochondria, the information transmitted is of maternal origin.

Cytoskeleton Network of stabilizing protein in the cytoplasm and cell membrane.

Cytosol The content of the cytoplasm of a cell, excluding cell organelles.

Dalton A unit of atomic mass, approximately equal to the mass of a hydrogen atom (1.66×10^{24} g).

Deficiency (Bridges 1917) Loss of a chromosomal segment resulting from faulty crossing-over, e.g., by unequal crossing-over or by crossing-over within an inversion (q.v.) or within a ring chromosome (q.v.). It arises at the same time as a complementary duplication (q.v.). This is referred to as duplication/deficiency.

Deletion (Painter and Muller 1929) Loss of part of or a whole chromosome or loss of DNA nucleotide bases.

Denaturation Reversible separation of a double-stranded nucleic acid molecule into single strands (see Renaturation).

Diakinesis (Haecker 1897) A stage during late prophase I of meiosis.

Dicentric (Darlington 1937) Refers to a structurally altered chromosome with two centromeres.

Dictyotene A stage of fetal oocyte development during which meiotic prophase is interrupted. In human females, oocytes attain the stage of dictyotene ~4 weeks before birth; further development of the oocytes is arrested until ovulation, at which time meiosis is continued.

Differentiation The process in which an unspecialized cell develops into a distinct, specialized cell.

Diploid (Strasburger 1905) Cells or organisms that have two homologous sets of chromosomes, one from the father (paternal) and one from the mother (maternal).

Diplotene A stage of prophase I of meiosis.

Direct repeat Repeated DNA sequences oriented in the same direction (see Inverted repeat).

Discordance The occurrence of a given trait or disease in only one member of a pair of twins (see Concordance).

Disomy, uniparental (UPD) Presence of two chromosomes of a pair from only one of the parents. One distinguishes UPD due to isodisomy, in which the chromosomes are identical, and heterodisomy, in which they differ.

Dispermy The penetration of a single ovum by two spermatozoa.

Divergence time The time elapsed since two populations have split from a common ancestor (see Coalescence time).

Dizygotic Twins derived from two different zygotes (fraternal twins), as opposed to monozygotic (identical) twins, derived from the same zygote.

D-loop Displacement loop; a DNA sequence that is formed in a part of opened DNA double helix, e.g., in mitochondrial DNA or in the telomere.

DNA (deoxyribonucleic acid) The molecule containing the primary genetic information in the form of a linear sequence of nucleotides in groups of threes (triplets; see codon).

DNase An enzyme that cleaves bonds in DNA.

DNA library A collection of cloned DNA molecules comprising the entire genome (genomic library) or of cDNA fragments obtained from mRNA produced by a particular cell type (cDNA library).

DNA chip or microarray A set of many thousands of different nucleotide sequences on a surface. Used to determine the gene expression pattern of thousands of genes simultaneously.

DNA methylation The presence of methyl groups at specific sites by changing cytosine to 5-methylcytosine.

DNA polymerase A DNA-synthesizing enzyme. To begin synthesis, it requires a primer of RNA or a complementary strand of DNA.

Domain A distinctive functional region of the tertiary structure of a protein.

Dominant (Mendel 1865) Refers to a genetic trait that can be observed in the heterozygous state. This depends in part on the accuracy of observation (see Codominant and Recessive).

Dominant negative A mutant allele that produces an undesirable effect resembling loss of function (see gain-of-function mutation).

Dorsal Refers to the back of an animal (for opposite, see ventral).

Dosage compensation (Muller 1948) Refers to mechanisms that balance a difference in activity of alleles.

Downstream The 3′ end of a gene.

Dynamic mutation A change in DNA characterized by expansion of nucleotide repeats.

Drift, genetic (Wright 1921) Random changes in gene frequency over generations in small populations. Under some conditions, an allele may disappear completely from a population or be fixed, i.e., present in all individuals.

Duplication (Bridges 1919) Addition of a chromosomal segment resulting from faulty crossing-over (see Deficiency). It may also refer to additional DNA nucleotide base pairs. Duplication of genes (gene duplication) played an important role in the evolution of eukaryotes.

Ectoderm One of the three primary cell layers of an embryo giving rise to epidermal tissues, the nervous system, and external sense organs (see Endoderm and Mesoderm).

Effector A protein exerting a specific effect.

Electrophoresis (Tiselius 1937) Separation of molecules by utilizing their different speeds of migration in an electrical field. As a support medium, substances in gel form such as starch, agarose, and acrylamide are used. Further molecular differences can be detected by modifications such as two-dimensional electrophoresis (electric field rotated 90 degrees for the second migration) or cessation of migration at the isoelectric point (isoelectric focusing).

Elongation Addition of amino acids to a polypeptide chain.

Elongation factor (EF) One of the proteins that associates with ribosomes while amino acids are added. EF in prokaryotes and eEF in eukaryotes.

Embryonic stem (ES) cells Undifferentiated pluripotent cells derived from an embryo.

Endocytosis Specific uptake of extracellular material at the cell surface. The material is surrounded by an invagination of the cell membrane, and this fuses to form a membrane-bound vesicle containing the material.

Endoderm The inner of the three primary cell layers of an embryo; gives rise to the gastrointestinal system and most of the respiratory tract (see Ectoderm and Mesoderm).

Endonuclease A heterogeneous group of enzymes that cleaves bonds between nucleotides of single-stranded or double-stranded DNA or RNA.

Endoplasmic reticulum A complex system of membranes within the cytoplasm.

Endoreduplication (Levan and Hauschka 1953) Chromosome replication during interphase without actual mitosis. Endoreduplicated chromosomes in metaphase consist of four chromatids lying next to each other, held together by two neighboring centromeres.

Enhancer (Baneriji 1981) A cis-acting regulatory DNA segment that contains binding sites for transcription factors. An enhancer is located at various distances from the promoter. It causes an increase in the rate of transcription (~10-fold).

Enzyme (Büchner 1897) A protein that catalyzes a biochemical reaction. Enzymes consist of a protein part (apoenzyme), responsible for the specificity, and a nonprotein part (coenzyme), needed for activity. Enzymes bind to their substrates, which become metabolically altered, or they combine with other substances during the train of the reaction. Most of the enzymatically catalyzed chemical reactions can be classified into one of six groups: (1) hydrolysis (cleavage with the addition of H_2O), by hydrolases; (2) transfer of a molecular group from a donor to a receptor molecule, by transferases; (3) oxidation and reduction, by oxidases and reductases (transfer of one or more electrons or hydrogen atoms from a molecule to be oxidized to another molecule that is to be reduced); (4) isomerization, by isomerases (rearranging the position of an atom or functional group within a molecule); (5) joining of two substrate molecules to form a new molecule, by ligases (synthetases); (6) nonhydrolytic cleavage with formation of a double bond on one or both of the two molecules formed, by lyases.

Epigenetics (Waddington 1942) The study of heritable changes in gene expression caused by mechanisms that do not alter the DNA sequence, often by DNA methylation. A portmanteau (blended word) of epigenesis and genetics.

Epigenome Refers to the overall epigenetic state of a genome.

Episome (Jacob and Wollman 1958) A DNA sequence, e.g., in plasmid (q.v.), that can exist either independently in the cytoplasm or as an integrated part of the genome of its bacterial host.

Epistasis (Bateson 1907) Nonreciprocal interaction of genes at the same gene locus (allelic) or at different gene loci (nonallelic) that alter the phenotypic expression of a gene.

Epitope The part of an antigen molecule that binds to an antibody.

EST (expressed sequence tag) A sequenced site from an expressed gene that "tags" a stretch of unsequenced cDNA next to it; used to map genes (see STS).

Eubacteria A major class of prokaryotes (see p. 26).

Euchromatin (Heitz 1928) Chromosome or chromosomal segment that stains less intensely than heterochromatin (q.v.). Euchromatin corresponds to the genetically active part of chromatin that is not fully condensed in the interphase nucleus.

Eukaryote (Chatton 1937) Cells in animals and plants that contain a nucleus and organelles in the cytoplasm (see Eubacteria and Prokaryote).

Euploid (Täckholm 1922) Refers to cells, tissues, or individuals with the complete normal chromosomal complement characteristic of that species (compare with Aneuploidy, Heteroploid, Polyploid).

Excision repair Repair of bulk lesions in DNA in which a stretch of nucleotides (~14 in

prokaryotes and ~30 in eukaryotes) is excised from the affected strand and replaced by the normal sequence (resynthesis).

Exocytosis Specific process by which nondiffusable particles are transported through the cell membrane to be discharged into the cellular environment.

Exome All exons in a genome.

Exon (Gilbert 1978) A segment of DNA that is represented in the mature mRNA of eukaryotes (compare with Intron).

Exonuclease An enzyme that cleaves nucleotide chains at their terminal bonds only, at either the 5′ or 3′ end (compare with Endonuclease).

Expression The observable effects of an active gene.

Expression vector A cloning vector containing DNA sequences that can be transcribed and translated.

Expressivity (Vogt 1926) Refers to the degree of phenotypic expression of a gene or genotype. Absence of expressivity is also called nonpenetrance.

Fibroblast A type of connective tissue cell. Can be propagated in culture flasks containing suitable medium (fibroblast cultures).

Fingerprint, genetic A characteristic pattern of small polymorphic fragments of DNA or proteins.

FISH (fluorescence *in-situ* hybridization) Identification of a DNA stretch on a surface (in situ) by fluorescent markers.

Fitness, biological Refers to the probability (between 0.0 and 1.0) that a gene will be transmitted to the next generation. For a given genotype and a given environment, the biological (or reproductive) fitness is determined by survival rate and fertility.

Fixation An allele that becomes permanently present in all members of a population.

Fluorophore A macromolecule bound to a fluorescent molecule with a defined light wavelength that allows recognition at the site of hybridization (see FISH).

Founder effect Presence of a particular allele in a population as a result of a mutation in a single ancestor.

Fragile site A specific site on a metaphase chromosome that appears to be broken because of a local difference in chromosome condensation. Inducible by culture medium conditions.

Frameshift mutation A mutation that alters the reading frame of a triplet in mRNA. As a consequence, an incorrect amino acid is inserted or a stop codon is created.

Gain-of-function A mutation that causes a new type of function, usually undesirable (see Dominant negative mutation).

Gamete (Strasburger 1877) A haploid germ cell, either a spermatozoon (male) or an ovum (female). In mammals, males are heterogametic (XY) and females homogametic (XX). In birds, females are heterogametic (ZW) and males homogametic (ZZ).

G-bands A type of banding pattern of metaphase chromosomes used for their identification.

Gene (Johannsen 1909) A hereditary factor that constitutes a single unit of genetic information. It corresponds to a segment of DNA that encodes the synthesis of a single polypeptide chain (compare with Cistron).

Gene amplification (Brown and David 1968) Selective production of multiple copies of a given gene without proportional increases of other genes.

Gene bank A collection of cloned DNA fragments that together represent the genome from which they are derived (see Gene library).

Gene cluster (Demerec and Hartman 1959) A group of two or more neighboring genes of similar function, e.g., the HLA system or the immunoglobulin genes.

Gene conversion (Winkler 1930; Lindgren 1953) Nonreciprocal transfer of genetic information. One gene serves as a sequence donor, remaining unaffected, while the other gene receives sequences and undergoes variation.

Gene desert A region along a chromosome that contains fewer genes than other regions.

Gene dosage Refers to the quantitative degree of expression of a gene.

Gene family A set of evolutionarily related genes by virtue of identity or great similarity of some of their coding sequences.

Gene flow (Berdsell 1950) Transfer of an allele from one population to another.

Gene frequency The frequency of a given allele at a given locus in a population (allele frequency).

Gene locus (Morgan et al. 1915) The position of a gene on a chromosome.

Gene map The position of gene loci on chromosomes. A physical map refers to the absolute position of gene loci, with their distance from each other expressed as the number of base pairs between them. A genetic map expresses the distance of genetically linked loci by their frequency of recombination.

Gene pool All genes present in a particular population.

Gene product The polypeptide or ribosomal RNA encoded by a gene (see Protein).

Genetic code The information contained in the triplets of DNA nucleotide bases used to incorporate a particular amino acid into a gene product.

Genetic distance The distance on a genetic map as defined by the frequency of recombination (recombination fraction), measured in centimorgans (cM).

Genetic marker A polymorphic genetic DNA sequence that can be used to distinguish the parental origin of alleles.

Genetics (Bateson 1906) The science of heredity and variation. The hereditary basis of organisms; derived from the Greek word genesis (origin).

Genome (Winkler 1920) All of the genetic material of a cell or of an individual.

Genome browser An electronic data bank providing a graphical representation of genes along a stretch of DNA (see UCSC Genome Bioinformatics at https://genome.ucsc.edu/).

Genomic disorders A group of disorders resulting from certain structural features of the human genome.

Genome scan A search with marker loci on all chromosomes for linkage with an unmapped locus.

Genome-wide association study (GWAS) A search for the sites of an association of disease susceptibility with polymorphic DNA markers (usually SNPs) in the whole genome.

Genomics The scientific field dealing with the structure and function of entire genomes.

Genotype (Johannsen 1909) All or a particular part of the genetic constitution of an individual or a cell (compare with Phenotype).

Germ cell A cell able to differentiate into gametes by meiosis (as opposed to somatic cells).

Germinal Refers to germ cells, as opposed to somatic cells.

Germline The cell lineage giving rise to germ cells.

Germline mosaic The presence of a proportion of germ cells in a gonad carrying a mutation.

G protein Guanine-nucleotide-binding proteins involved in signal transduction.

G6PD Glucose-6-phosphate dehydrogenase.

Growth factor A protein, usually a small peptide, acting as a ligand that activates a receptor.

Gyrase A topoisomerase that unwinds DNA.

Haldane's rule Hybrid sterility or inviability preferentially affecting the heterogametic sex.

Haploid (Strasburger 1905) Refers to cells or individuals with a single chromosome complement; gametes are haploid.

Haploinsufficiency Refers to a diploid gene that does not exert normal function in the haploid state (e.g., following loss of the homologous gene locus).

Haplotype (Ceppellini et al. 1967) A series of alleles at two or more closely linked gene loci

on the same chromosome, e.g., in the HLA system (q.v.).

Hardy–Weinberg equilibrium The distribution of frequencies of genotypes in a population.

Heat map A graphical display of gene expression in array data. A row represents a gene, a column represents a sample of genes. The color of each gene represents the gene expression (high activity usually in red, low in blue).

Heavy strand Refers to differences in density of DNA that result from differences in A and G, and C and T bases, in contrast to the light strand.

Helicase An enzyme that unwinds and separates the two strands of the DNA double helix by breaking the hydrogen bonds during transcription or repair.

Helix–loop–helix A structural motif in DNA-binding proteins, such as some transcription factors.

Hemizygous Refers to genes and gene loci that are present in only one copy in an individual, e.g., on the single X chromosome in male cells (XY) or because a homologous locus has been lost.

Heritability (Lush 1950; Falconer 1960) The ratio of additive genetic variance to the total phenotypic variance. Phenotypic variance is the result of the interaction of genetic and nongenetic factors in a population.

Heterochromatin (Heitz 1928) A chromosome or chromosomal segment that remains darkly stained in interphase, early prophase, and late telophase because it remains condensed, as all chromosomal material is in metaphase. This contrasts with euchromatin, which becomes invisible during interphase. Heterochromatin corresponds to chromosomes or chromosome segments showing little or no genetic activity. Constitutive and facultative heterochromatin can be distinguished. An example of facultative heterochromatin is the heterochromatic X chromosome resulting from inactivation of one X chromosome in somatic cells of female mammals. An example of constitutive heterochromatin is the centric heterochromatin at centromeres that can be demonstrated as C bands.

Heterodisomy Presence of two homologous chromosomes from one parent only (compare with Isodisomy and UPD).

Heteroduplex Refers to a region of a double-stranded DNA or DNA/RNA molecule with noncomplementary strands that originated from different duplex DNA molecules.

Heterogametic (Wilson 1910) Refers to the two different types of gametes (q.v.), e.g., X and Y in (male) mammals or Z and W in female birds.

Heterogeneity, genetic (Harris 1953; Fraser 1956) An apparently uniform phenotype caused by two or more different genotypes.

Heterokaryon (Ephrussi and Weiss 1965; Harris and Watkins 1965; Okada and Murayama 1965) A cell having two or more nuclei with different genomes.

Heteroplasmy Differences in the presence or absence of mutations in mitochondria within a single cell.

Heteroploid (Winkler 1916) Refers to cells or individuals with an abnormal number of chromosomes.

Heterosis (Shull 1911) Increased reproductive fitness of heterozygous genotypes compared with the parental homozygous genotypes, in plants and animals.

Heterozygous (Bateson and Saunders 1902) Having two different alleles at a given gene locus (compare with Homozygous).

Hfr cell A bacterium that possesses DNA sequences that lead to a high frequency of DNA transfer at conjugation.

HGPRT Hypoxanthine-guanine-phosphoribosyl transferase. An enzyme in purine metabolism that is inactive in Lesch–Nyhan syndrome.

Histocompatibility Tissue compatibility. Determined by the major histocompatibility complex MHC (see HLA).

Histone (Kossel 1884) Chromosome-associated protein of the nucleosome. Histones H2A, H2B, H3, and H4 form a nucleosome (q.v.).

Histone code A pattern of chemical modifications in histones (e.g., methylation, acetylation) that are related to the regulation of gene activities.

HLA (Dausset and Terasaki 1954) Human leukocyte antigen system A. HLA is also said by some to refer to Los Angeles, where Terasaki made essential discoveries.

HMG proteins High-mobility group proteins that exhibit high mobility during electrophoresis. Nuclear proteins involved in different DNA interactions.

Hogness box A nucleotide sequence (5′-TATAAA-3′) located ~25 bp upstream (in 5′ direction) of the promoter region in eukaryotic genes (usually called TATA box).

Holoprosencephaly A wide range of congenital malformations involving the midline embryonic forebrain, ranging from a single eye (cyclopia) without nose formation at the most severe end of the spectrum to a single incisor tooth and a flat face with closely set eyes (hypotelorism) as a result of mutations in several genes.

Homeobox A highly conserved DNA segment in homeotic genes.

Homeosis Transformation of one body part into another.

Homeotic gene One of the developmental genes in *Drosophila* that can lead to the replacement of one body part by another by mutation.

Homologous Refers to a chromosome or gene locus of similar maternal or paternal origin.

Homozygosity mapping Mapping genes by identifying chromosomal regions that are homozygous as a result of identity by descent from a common ancestor in consanguineous matings (see Identity by descent, IBD).

Homozygous (Bateson and Saunders 1902) Having identical alleles at a given gene locus.

Hormone (Bayliss and Starling 1904) An organic compound able to induce a specific response in target cells. (Derived from the Greek “to spur on.”)

Hotspot A site on a chromosome where recombination frequently and preferentially occurs.

Housekeeping gene A gene carrying information about basic functions in a cell.

Hox genes Clusters of mammalian genes containing homeobox sequences. They are important in embryonic development.

Hybridization In genetics, the fusion of two single complementary DNA strands (DNA–DNA hybridization) DNA and RNA strands (DNA–RNA hybridization), or the in vitro fusion of cultured cells of different species (cell hybridization).

Hybridoma A clone of hybrid cells.

Hydrogen bond A noncovalent weak chemical bond between an electronegative atom (usually oxygen or nitrogen) and a hydrogen atom; important in stabilizing the three-dimensional structure of proteins or base pairing in nucleic acids.

Hypomorph An allele that produces a reduced amount or activity of a gene product.

Identity by descent (IBD) Refers to homozygous alleles at one gene locus that are identical because they are inherited from a common ancestor (see Consanguinity).

Immunoglobulin An antigen-binding molecule.

Imprinting, genomic Different expression of a gene depending on the parental origin.

Inbreeding coefficient (Wright 1929) Measure of the probability that two alleles at a gene locus of an individual are identical by descent, i.e., that they are copies of a single allele of an ancestor common to both parents (IBD, identity-by-descent). Also, the proportion of loci at which the individual is homozygous.

Incidence A measure of the rate of occurrence of a disease in a population in a specified period of time (compare with Prevalence).

Indel Insertion/deletion variant in DNA.

Index patient The patient who first calls attention to a family and leads to ascertainment.

Inducer A molecule that induces the expression of a gene.

Initiation factor A protein that associates with the small subunit of a ribosome when protein synthesis begins (IF in prokaryotes, eIF in eukaryotes).

Insertion Insertion of chromosomal material of nonhomologous origin into a chromosome without reciprocal translocation (q.v.).

Insertion sequence (IS) A small bacterial transposon carrying genes for its own transposition (q.v.).

In silico A process taking place within a computer; e.g., analysis of biological data.

Intercalating agent A chemical compound that can occupy a space between two adjacent base pairs in DNA.

In-situ hybridization Hybridization of complementary single-stranded DNA or RNA in the original or natural position (see FISH).

Interphase The period of the cell cycle between two cell divisions (see Mitosis).

Intron (Gilbert 1978) A segment of noncoding DNA within a gene (compare with Exon). It is transcribed, but removed from the primary RNA transcript before translation.

Inversion (Sturtevant 1926) Structural alteration of a chromosome through a break at two sites with reversal of direction of the intermediate segment and reattachment. A pericentric inversion includes the centromere in the inverted segment. A paracentric inversion does not involve the centromere.

Inverted repeat Two identical, oppositely oriented copies of the same DNA sequence. They are a characteristic feature of retroviruses.

In vitro A biological process taking place outside a living organism or in an artificial environment in the laboratory.

In vivo Within a living organism.

iPS cells (iPSCs) (Yamanaka 2006) Induced pluripotent stem cells. Adult cells that have been reprogrammed to express genes (Oct3, Oct4, SOX2, c-Myc, KLF4) that are active in pluripotent stem cells only, such as embryonic stem cells.

Isochromatid break A break in both chromatids at the same site.

Isochromosome (Darlington 1940) A chromosome composed of two identical arms connected by the centromere, e.g., two long or two short arms of an X chromosome. Implies duplication of the doubled arm and deficiency of the absent arm. An isochromosome may have one or two centromeres.

Isodisomy Presence of two identical chromosomes from one of the parents (compare with Heterodisomy).

Isolate, genetic (Waklund 1928) A physically or socially isolated population that has not interbred with individuals outside of that population (no panmixia).

Isotype Closely related chains of immunoglobulins.

Isozyme or isoenzyme (Markert and Möller 1959; Vesell 1959) One of multiple distinguishable polymorphic variants of enzymes of similar function in the same organism. Isoenzymes are a biochemical expression of genetic polymorphism.

IVS Intervening sequences. DNA sequences in an intron.

Jewish genetic diseases As a result of relative isolation, several autosomal recessive diseases occur more frequently in Jewish populations (see Jewish Genetic Disease Consortium at: http://www.jewishgeneticdiseases.org/; Accessed July 6, 2017).

Karyogram The visual display of all chromosomes of a cell arranged in pairs of homologous chromosomes.

Karyotype (Levitsky 1924) The complete chromosome set of a cell, an individual, or a species.

Kilobase (kb) 1,000 base pairs (bp).

Kinetochore Chromosomal structure at the centromere, where the spindle fibers attach during cell division.

Knockout (of a gene) Intentional inactivation of a gene in an experimental organism to obtain information about its function (same as targeted gene disruption).

Lagging strand of DNA The new strand of DNA replicated from the 3′ to 5′ strand. It is synthesized in short fragments in the 5′ to 3′ direction (Okazaki's fragments), and these are subsequently joined together.

Lamins Intermediate filament proteins forming a fibrous network, the nuclear lamina on the inner surface of the nuclear envelope.

Lampbrush chromosome (Rückert 1892) A special type of chromosome found in the primary oocytes of many vertebrates and invertebrates during the diplotene stage of meiotic division and in *Drosophila* spermatocytes. The chromosomes show numerous lateral loops of DNA that are accompanied by RNA and protein synthesis.

Lariat An intermediate form of RNA during splicing when a circular structure with a tail is formed by a 5′ to 3′ bond.

Leader sequence A short N-terminal sequence of a protein that is required for directing the protein to its target.

Leaky mutant A mutation causing only partial loss of function.

Leptotene A stage of meiosis (q.v.).

Lethal equivalent (Morton et al. 1956) A gene or combination of genes that in the homozygous state is lethal to 100% of individuals. This may refer to a gene that is lethal in the homozygous state, to two different genes each, with 50% lethality, or to three different genes, each with 33% lethality, etc. It is assumed that each individual carries about five or six lethal equivalents.

Lethal factor (Bauer 1908; Hadorn 1959) An abnormality of the genome that leads to death in utero, e.g., numerous chromosomal anomalies.

Leucine zipper A specific DNA-binding protein that resembles a zipper and serves as a transcription factor.

Library See DNA library.

Ligand A molecule that can bind to a receptor and induce a signal in the cell.

Ligase An enzyme that closes a gap in a DNA strand.

LINE (long interspersed nuclear element) Long interspersed repetitive DNA sequences.

Linkage disequilibrium (Kimura 1956) Nonrandom association of alleles at closely linked gene loci that deviates from their individual frequencies as predicted by the Hardy–Weinberg equilibrium.

Linkage, genetic (Morgan 1910) Localization of gene loci on the same chromosome close enough to cause deviation from independent segregation.

Linkage group Gene loci on the same chromosome that are so close together that they usually are inherited together without recombination.

Linker DNA A synthetic DNA double-stranded fragment that carries the recognition site for a restriction enzyme and that can bind two DNA fragments. Also, the stretch of DNA between two nucleosomes.

Locus Same as gene locus.

Loss-of-function mutation A mutation that causes total or partial loss of function of a gene product (see Gain-of-function mutation).

Loss of heterozygosity (LOH) Hemizygosity or homozygosity in a tumor resulting from loss of a heterozygous gene locus or chromosome.

Long terminal repeat (LTR) A repeat DNA sequence of up to 600 bp flanking the coding regions of retroviral DNA and viral transposons.

Lymphocyte Cell of the immune system, of one of two general types: B lymphocytes from the bone marrow and thymus-derived T lymphocytes.

Lysogeny The ability of a phage to integrate into the bacterial chromosome.

Lysosome (de Duve 1955) Small cytoplasmic organelle containing hydrolytic enzymes.

Lytic infection A phage infecting a bacterium and causing lysis.

Map distance Distance between gene loci, expressed in either physical terms (number of base pairs, e.g., kb [1,000 bp] or Mb [million bp]) or genetic terms (recombination frequency, expressed as cM, centimorgan. One cM corresponds to 1%).

Mapping Various methods to determine the position of a gene on a chromosome (physical map) or its relative distance to other gene loci and their order (genetic map).

Marker, genetic An allele used to recognize a particular genotype.

Megabase (Mb) 1 million base pairs (bp).

Meiosis (Strasburger 1884) The special division of a germ cell nucleus that leads to reduction of the chromosome complement from the diploid to the haploid state. Prophase of the first meiotic division is especially important and consists of the following stages: leptotene, zygotene, pachytene, diplotene, diakinesis.

Mendelian inheritance (Castle 1906) Inheritance according to the laws of Mendel as opposed to extrachromosomal inheritance, under the control of cytoplasmic hereditary factors (mitochondrial DNA).

Metabolic cooperation (Subak-Sharpe et al. 1969) Correction of a phenotype in cells in culture by contact with normal cells or cell products.

Metacentric Refers to chromosomes that are divided by the centromere into two arms of approximately the same length.

Metaphase (Strasburger 1884) Stage of mitosis in which the contracted chromosomes are readily visible.

MHC (major histocompatibility complex) (Thorsby 1974) The principal histocompatibility system, consisting of class I and class III antigen genes of the HLA system, including the class II genes.

Microfilament Cytoskeletal fiber of ~7-nm diameter that results from polymerization of monoglobar (G) actin (actin filament).

microRNA (miRNA) Small RNA molecules of 21–23 nucleotides that associate with multiple proteins in an RNA-induced silencing complex (RISC; see RNAi).

Microsatellite DNA Small tandem repeats of ~1–6 bp (e.g., CACACACA), usually less than 0.1 kb in total length; also referred to as short tandem repeats (STRs).

Microsatellite instability Instability of microsatellites in some forms of cancer, especially in the colon.

Mismatch repair A DNA repair mechanism to repair improperly paired DNA bases.

Missense mutation A mutation that alters a codon to one for a different amino acid (see Nonsense mutation).

Mitochondrion (plural: mitochondria) Cell organelles containing DNA.

Mitosis (Flemming 1882) Nuclear division during the division of somatic cells, consisting of prophase, metaphase, anaphase, and telophase.

Mitosis index (Minot 1908) The proportion of cells present that are undergoing mitosis.

Mixoploidy (Nemec 1910; Hamerton 1971) A tissue or individual containing cells with different numbers of chromosomes (chromosomal mosaic).

MLC (Bach, Hischhorn, Lowenstein 1964) Mixed lymphocyte culture. MLC is a test for differences in HLA-D phenotypes.

Mobile genetic element A DNA sequence that can change its position (see Transposon).

Modal number (White 1945) The number of chromosomes of an individual or a cell.

Monoclonal antibody An antibody representing a single antigen specificity, produced from a single progenitor cell.

Monolayer (Abercrombie and Heaysman 1957) The single-layered sheet of cultured diploid cells on the bottom of a culture flask.

Monosomy (Blakeslee 1921) Absence of a chromosome in an otherwise diploid chromosomal complement (see Trisomy).

Monozygotic Twins with identical sets of nuclear genes (compare with Dizygotic).

Morphogen A protein present in embryonic tissues in a concentration gradient that induces a developmental process.

Mosaic Tissue or individuals made up of genetically different cells, as a rule of the same zygotic origin (compare with Chimera).

Mousterian Middle Stone Age tools mainly associated with Neanderthals (compare with Aurignacian).

mRNA (Brenner et al. 1961; Jacob and Monod, 1961) Messenger RNA.

mtDNA Mitochondrial DNA.

Multigene family A group of genes related by their common evolution.

Mutagen A chemical or physical agent that can induce a mutation.

Mutation (de Vries 1901) A heritable structural and functional alteration of the DNA sequence. Different types include point mutations from exchange, loss, or insertion of base pairs within a gene (see Missense mutation and Nonsense mutation).

Mutation nomenclature Mutations are designated according to an international nomenclature system by the Human Genome Variation Society (HGVS). It indicates the type of reference sequence used by a letter prefix: "g." for genomic reference sequence, "c." for coding sequence, "n" for noncoding, "r." for RNA, "p." for protein. All variants should be described in relation to an accepted reference sequence. Substitutions at the DNA level are designated by the sign ">" and deletions by "del" after the deleted interval and insertions by "ins" followed by the inserted nucleotides. Thus, 1994G>T denotes a change of G (guanine) to T (thymine) at nucleotide (nt) position 1994. 1994–1996delTTC denotes a deletion of two T and one C (cytosine). 1994–1995insAG denotes an insertion of an A (adenine) and G between nt position 1994 and 1995. Intron changes are designated by "IVS" followed by the intron number or the cDNA position. g.10G>C describes a change of G to C at nt position 10 in relation to genomic DNA; c.10g>c describes this change in cDNA and m.10G>C in mitochondrial DNA. Mutations designated at the protein level indicate the deduced consequence preceded by "p" e.g., p. R25S designates a change of arginine (R) to serine (S) at amino acid position 25. Deletions and insertions at the protein level are designated by "del" or "ins" after the positions of the amino acids involved: R98–S101del means a deletion between R at position 98 and S at 101. R98–S99insR means an insertion of R between R at 98 and S at 99. Note that at the protein level the deduced consequence is described, but not the nature of the mutation. (For details, see den Dunnen JT, Antonarakis SE. Mutation nomenclature. Hum Mutat 2000; 15: 7–12; Human Mutation Special Issue June 2016; 37: 499–615; and Human Sequence Variant Nomenclature at http://varnomen.hgvs.org).

Mutation rate The frequency of a mutation per locus per individual per generation. The average mutation rate in man is estimated to be ~2.5×10^{-8} (Nachman MW, Crowell SL, Genetics 2000; 156: 297–304).

Myosin A class of motor proteins.

Necrosis Cell death resulting from tissue damage.

Neural crest A structure parallel to the dorsal neural tube containing a network of migratory cell populations that give rise to different cell types and anatomical structures.

Neurotransmitter Extracellular signaling molecules at the nerve–muscle junction.

N-linked oligosaccharides Branched oligosaccharide chains attached to the amino group side chain of asparagine in glycoproteins.

Nonallelic homologous recombination (NAHR) Recombination between misaligned DNA repeats on the same chromosome or on sister chromatids. The consequences are deletions or duplications (see Genomic disorders).

Noncoding RNA (ncRNA) RNA not involved in transcription and translation, but with different functions (see Coding RNA).

Noncovalent bond A (weak) chemical bond between an electron-negative atom (usually oxygen or nitrogen) and a hydrogen atom (see hydrogen bond), but not involving sharing of electrons.

Nondisjunction (Bridges 1912) Faulty distribution of homologous chromosomes at meiosis. In mitotic nondisjunction, the distribution error occurs during mitosis.

Nonpolar Molecules lacking a net electric charge or having an asymmetric distribution of positive and negative charges; generally insoluble in water.

Nonsense codon A codon that does not have a normal tRNA molecule. Any of the three triplets (stop codons) that terminate translation (UAG, UAA, UGA).

Nonsense-mediated mRNA decay A mechanism to degrade mRNA molecules containing a premature stop codon (usually <50 nucleotides upstream of the last splice junction) before translation into a polypeptide.

Nonsense mutation A mutation that results in lack of any genetic information, e.g., a stop codon (see Missense mutation).

Northern blot Transfer of RNA molecules to a membrane by a procedure similar to that of a Southern blot (q.v.). A word play on Southern blot (see Western blot).

Nucleic acid A molecule such as DNA and RNA that can store genetic information.

Nucleolus The site in the nucleus where ribosomal RNA is transcribed.

Nucleolus organizer (NOR) The satellite stalks of human acrocentric chromosomes 13, 14, 15, 21, and 22 containing ribosomal RNA that can be specifically stained with silver.

Nucleoside Compound of a purine or a pyrimidine base with a sugar (ribose or deoxyribose) (compare with Nucleotide).

Nucleosome (Navashin 1912; Kornberg 1974) A subunit of chromatin consisting of DNA wound around histone proteins in a defined spatial configuration.

Nucleotide Single monomeric building block of a polynucleotide chain that makes up nucleic acid. A nucleotide is a phosphate ester consisting of a purine or a pyrimidine base, a sugar (ribose or deoxyribose as a pentose), and a phosphate group.

Ochre codon The stop codon UAA (see Amber codon, and Stop codon UAG).

Okazaki's fragments Short (<1,000 nucleotides) single-stranded DNA fragments that are synthesized on the lagging strand of DNA during replication (q.v.). They are rapidly joined by DNA ligase into a continuous new DNA strand.

O-linked oligosaccharides Attachment to the hydroxyl groups of the side chains of serine and threonine in glycoproteins (see N-linked).

Oncogene (Huebner and Todaro 1969) Originally a DNA sequence of viral origin that can lead to malignant transformation of a eukaryotic cell after being integrated into the cellular DNA. Now used to designate a gene involved in control of cell proliferation (formerly proto-oncogene).

Open reading frame (ORF) A DNA sequence of variable length that does not contain a stop codon within any of the triplet reading frames. Any ORF beginning with a start codon and extending for 100 or more codons probably encodes a protein, and is therefore meaningful because it can be translated.

Operator (Jacob and Monod 1959) The recognition site of an operon at which the negative control of genetic transcription takes place by binding to a repressor.

Operon (Jacob et al. 1960) In prokaryotes, a group of functionally and structurally related genes that are regulated together.

Organelle A membrane-enclosed subcellular structure in a eukaryotic cell (see Cell organelle).

Origin of replication (ORI) Site of start of DNA replication.

Ortholog A homologous DNA sequence or a gene that has evolved by speciation from a common ancestor between related species, e.g., the α- and β-globin genes (see paralog).

Pachytene (de Winiwarter 1900) Stage of prophase meiosis I.

Palindrome (Wilson and Thomas 1974) A sequence of DNA nucleotides that reads the same in the 5′ to 3′ direction as in the 3′ to 5′ direction. This occurs in recognition sequences of restriction enzymes, e.g., 5′-AATG-3′ in one strand and 3′-GTAA-5′ in the other.

Panmixia (Weismann 1895) Pairing system with random partner selection, as opposed to assortative mating.

Paracrine Refers to secreted signaling molecules acting over a short distance on neighboring cells.

Paralog A DNA sequence or a gene that has evolved by duplication from a common ancestor within a species, e.g., the two α-globin gene loci in humans (see ortholog).

Parasexual (Pontocorvo 1954) Refers to genetic recombination by nonsexual means, e.g., by hybridization of cultured cells (see Hybridization).

PCR (polymerase chain reaction) (Mullis 1985) Technique for in vitro propagation (amplification) of a given DNA sequence. It is a repetitive thermal cyclic process consisting of

denaturation of genomic DNA of the sequence of interest, annealing the DNA to appropriate oligonucleotide primers, and replication of the DNA segment complementary to the primer.

Penetrance (Vogt 1926) The frequency or probability of expression of an allele (compare with Expressivity) and become visible in the phenotype.

Peptide A compound of two or more amino acids joined by peptide bonds.

Phage Abbreviation for bacteriophage.

Phagocytosis Refers to cells that incorporate foreign cells, such as bacteria.

Pharmacogenetics (Vogel 1958; Vesell 1959) Study of the individual influence of certain genes on the metabolism and function of pharmaceutical substances (see p. 272).

Pharmacogenomics Study of genetic variation in response to pharmaceutical substances on a whole genome basis.

Phenocopy (Goldschmidt 1935) A nonhereditary phenotype that resembles a genetically determined phenotype.

Phenotype (Johannsen 1909) The observable effect of one or more genes on an individual or a cell.

Pheromone A signaling molecule that can alter the behavior or gene expression of other individuals of the same species.

Phosphodiester bond The chemical bond linking adjacent nucleotides of DNA or RNA.

Phytohemagglutinin (PHA) A protein substance obtained from kidney beans (*Phaseolus vulgaris*). Nowell (1960) discovered its ability to induce blastic transformation (see Transformation) and cell division in lymphocytes. This is the basis of phytohemagglutinin-stimulated lymphocyte cultures for chromosomal analysis.

Plasmid (Lederberg 1952) Autonomously replicating circular DNA structures found in bacteria. Although they are usually separate from the actual genome, they may become integrated into the host chromosome.

Plastids Any of several types of organelles found in plant cells, e.g., chloroplasts.

Pleiotropy (Plate 1910) Expression of a gene with multiple, seemingly unrelated phenotypic features.

Point mutation Alteration of the genetic code within a single codon. The possible types are the exchange (substitution) of a base: a pyrimidine for another pyrimidine (or a purine for another purine), i.e., a *transition* (Frese 1959), e.g., thymine for cytosine (or adenine for guanine); or the exchange of a pyrimidine for a purine or vice versa: *transversion*, i.e., thymine by adenine or vice versa. In addition to the two types of exchange, a point mutation may be due to the insertion of a nucleotide base or the deletion of one or several base pairs.

Polar Refers to molecules with a net electric charge or a symmetric distribution of positive and negative charges.

Polar body (Robin 1862) An involutional cell arising during oogenesis that does not develop further as an oocyte.

Polyadenylation The addition of multiple adenine residues at the 3′ end of eukaryotic mRNA after transcription.

Polycistronic messenger mRNA including coding regions from more than one gene (in prokaryotes).

Polygenic (Plate 1913; Mather 1941) Refers to traits that are based on several or numerous genes whose effects cannot be individually determined. The term multigenic is sometimes used instead.

Polymerases Enzymes that catalyze the polymerization of new DNA or RNA nucleotides from a template as in transcription and DNA replication.

Polymorphism, genetic (Ford 1940) Existence of more than one variant of a normal allele at a gene locus. By definition, frequency must exceed 1%. Polymorphisms exist at several levels, i.e., sequence of DNA, amino acids, certain structures of chromosomes, or phenotypic traits.

Polypeptide Linear polymer of amino acids, which form a protein.

Polyploid (Strasburger 1910) Refers to cells, tissues, or individuals having more than two

copies of the haploid genome, e.g., three (triploid) or four (tetraploid).

Polytene chromosome (Koltzoff 1934; Bauer 1935) refers to a special type of chromosome resulting from repeated endoreduplication of a single chromosome. Giant chromosomes arise in this manner (compare with Chromosome).

Population (Johannsen 1903) Individuals of a species that interbreed and constitute a common gene pool (compare with race).

Position effect (Sturtevant 1925) An effect on the expression of a gene depending on its location on a chromosome (Sturtevant AH. The effects of unequal crossing over at the bar locus in *Drosophila*. Genetics 1925; 10: 117–147).

Positional cloning Identification of a gene based on prior information of the position of its locus.

Premutation A state leading to a mutation (see dynamic mutations repeat expansion disorders.

Premature chromosome condensation (Johnson and Rao 1970) Induction of chromosomal condensation in an interphase nucleus after fusion with a cell in mitosis. Condensed S-phase chromosomes appear pulverized (so-called chromosome pulverization).

Prevalence The total number of individuals with a given disease in a population at a given time (compare with Incidence).

Pribnow box Part of a promoter (TATAAT sequence 10 bp upstream of the gene) in prokaryotes.

Primary transcript The RNA transcribed from a eukaryotic gene before processing (splicing, addition of the cap, and polyadenylation).

Primer A short DNA or RNA oligonucleotide of ~15–25 bases that specifically binds to a target sequence and initiates synthesis of a complementary strand.

Primordial germ cells (PGCs) Cells in the gonads that give rise to germ cells in the ovary or testis.

Prion Proteinaceous infectious particles that cause degenerative disorders of the central nervous system.

Proband The individual who first calls attention to a family with a given genetic disorder (same as propositus or index patient).

Probe A defined single-stranded DNA or RNA fragment, labeled for detection with a fluorophore, to identify complementary sequences by specific hybridization.

Prokaryote Single-celled microorganisms without a cell nucleus or intracellular organelles, such as bacteria and archaea (compare eukaryote).

Promoter A defined DNA region at the 5′ end of a gene that binds to transcription factors and RNA polymerase during the initiation of transcription. The ~10 sequence is the consensus sequence TATAATG ~10 bp upstream of a prokaryotic gene (Pribnow box).

Prophage A viral (phage) genome integrated into the bacterial (host) genome.

Prophase An early stage of mitosis or meiosis.

Propositus, proband The individual in a pedigree that has brought a family to attention for genetic studies.

Protein A molecule with a distinct biological function, consisting of one or more polypeptide subunits each with a defined amino acid sequence folded into a specific three-dimensional structure.

Proteome The complete set of all protein-encoding genes or all proteins produced.

Proteomics The study of all proteins in a cell or an organism.

Proto-oncogene A normal gene that can be changed into a tumor-promoting oncogene by a mutation.

Prototrophy Refers to cells or cell lines that do not require a special nutrient added to the culture medium (compare with Auxotrophy).

Provirus Duplex DNA derived from an RNA retrovirus and incorporated into a eukaryotic genome.

Pseudoautosomal region (PAR) Regions at the end of the X- and the Y-chromosome that participate in recombination as in autosomes.

Pseudogene DNA sequences that closely resemble a gene but have lost function due to

structural rearrangements, such as stop codon, deletion, insertion, or other structural changes. A processed pseudogene consists of DNA sequences that resemble the mRNA copy of the parent gene, i.e., it does not contain introns.

Pseudohermaphroditism A condition in which an individual has the gonads of one sex and phenotypic features of the opposite sex.

qPCR Quantitative PCR.

Quadriradial figure The configuration assumed when homologous segments of chromosomes involved in a reciprocal translocation pair at meiosis. Rarely, such a figure may occur during mitosis.

Race A population (q.v.) that differs from another population within the same species in the frequency of some of its gene alleles (Dunn LC. Heredity and Evolution in Human Populations. Cambridge, MA: Harvard University Press; 1967). Accordingly, the concept of race is flexible and relative, defined in relation to the evolutionary process. The term race erroneously divides humans into distinct categories defined by external phenotypic features such as skin color that are the result of adaptation to climate conditions during evolution. It is often associated with discriminatory reactions for which no genetic basis exists. The classification of individuals according to race is uncertain and of dubious biological value.

Reading frame A sequence of DNA nucleotides that can be read in triplets in the genetic code (compare with Open reading frame).

Receptor A transmembrane or intracellular protein involved in transmission of a cell signal.

Recessive (Mendel 1866) Refers to the genetic effect of an allele (q.v.) at a gene locus that is manifest as phenotype in the homozygous state only (see dominant).

Reciprocal translocation Mutual exchange of chromosome parts.

Recombinant DNA An artificial DNA molecule constructed of fragments of different origins. Widely used in molecular genetic analysis.

Recombination (Bridges and Morgan 1923) The formation of new combinations of chromosomes and genes as a result of crossing-over between homologous chromosomes during meiosis. Recombination is the main source of genetic variation.

Recombination frequency Frequency of recombination between two or more gene loci. Expressed as the recombination fraction theta (Θ) value. A Θ of 0.01 (1% recombination frequency) corresponds to 1 centimorgan (cM).

Regulatory gene A gene encoding a protein that regulates other genes.

Renaturation of DNA Combining of complementary single strands of DNA to form double-stranded DNA (compare with Denaturation).

Repair (Muller 1954) Correction of structural and functional DNA damage.

Replication Identical duplication of the DNA double helix prior to cell division.

Replication fork The unwound region of the DNA double helix in which replication takes place.

Replicon (Huberman and Riggs 1968) An individual unit of DNA in eukaryotic DNA that is capable of replication.

Reporter gene A gene used to analyze another gene, especially in the regulatory region of the latter.

Repressor A protein that suppresses gene function.

Repulsion (Bateson et al. 1905) A term to indicate that the mutant alleles of neighboring heterozygous gene loci lie on opposite chromosomes, i.e., in trans configuration (see cis/trans).

Resistance factor A gene in plasmids causing antibiotic resistance.

Restriction enzyme, or restriction endonuclease (Meselson and Yuan 1968) An endonuclease that cleaves DNA at a specific base sequence (restriction site or recognition sequence).

Restriction fragment length polymorphism (RFLP) A variant DNA sequence (polymorphism) that creates or abolishes a restriction site. This can be used to identify the parental

origin of a given DNA fragment and used in diagnostics.

Restriction map A segment of DNA characterized by a particular pattern of absence or presence of a restriction site.

Restriction site A particular sequence of nucleotide bases in DNA at which a particular restriction enzyme cleaves a DNA molecule.

Retrotransposon A mobile DNA sequence that can insert itself at a different position by using reverse transcriptase (see Transposon).

Retrovirus A virus with a genome consisting of RNA that replicates by copying its RNA genome into cDNA prior to integration into a chromosome of the host cell.

Reverse transcriptase An enzyme complex present in RNA viruses that can synthesize DNA from an RNA template.

RFLP Restriction fragment length polymorphism.

Rho factor A protein involved in termination of transcription in Escherichia coli.

Ribonuclease (RNAase) Enzymes (nucleases) that can cleave RNA.

Ribosome (Dintzis et al. 1958; Roberts 1958) Complex molecular structure in prokaryotic and eukaryotic cells consisting of specific proteins and ribosomal RNA in different subunits. The translation of genetic information occurs in ribosomes.

Ring chromosome A circular chromosome. In prokaryotes the normal chromosome usually is ring shaped. In mammals a ring chromosome represents a structural anomaly and implies that chromosomal material has been lost.

RNA (ribonucleic acid) A polynucleotide with a structure similar to that of DNA except that the sugar is ribose instead of deoxyribose.

RNA editing A change of RNA sequence following transcription; a mechanism of gene regulation.

RNAi (RNA interference) Inhibition of transcription by antisense RNA (see p. 186).

RNA polymerase An enzyme that synthesizes RNA from a DNA template.

RNA silencing Downregulation or suppression of a gene by antisense RNA. The ability of double-stranded RNA to suppress a gene.

RNA splicing The processing of the primary transcript in mRNA (in eukaryotes).

rRNA Ribosomal RNA. Any of many large RNA molecules in ribosomes.

RTK Receptor tyrosine kinase; membrane bound proteins (G proteins) involved in signal transduction.

Satellite (Navashin 1912) Small mass of chromosomal material attached to the short arm of an acrocentric chromosome (q.v.) by a constricted appendage or stalk. It is involved in the organization of the nucleolus (see Nucleolus organizer).

Satellite DNA (sDNA) (Kit 1961; Sueoka 1961; Britten and Kohne 1968) DNA that forms additional minor band in density gradient centrifugation because it contains either heavier (GC-rich) or lighter (AT-rich) than the main DNA (see DNA). Not to be confused with the satellite regions of acrocentric chromosomes. sDNA can be separated from the main DNA by density gradient centrifugation in cesium chloride, after which it appears as one or several bands (satellites) separated from that of the main body of DNA. In eukaryotes, light (AT-rich) and heavy (GC-rich) satellite DNA can be distinguished. Microsatellites are small (2–10) tandem repeats of DNA nucleotides. Minisatellites are tandem repeats of ~20–100 bp; classical satellite DNA consists of large repeats of 100–6,500 bp.

SCE (sister chromatid exchange) (Taylor 1958) An exchange between the two chromatids of a metaphase chromosome. After two replication cycles in a cell culture in the presence of a halogenated base analogue (e.g., 5-bromodeoxyuridine), both DNA strands of one chromatid will be substituted with the halogenated base analogue, whereas only one DNA strand of the other chromatid will be substituted. As a result, the two chromatids differ in staining intensity and it is possible to determine where somatic crossing-over of the two chromatids has occurred. An increased rate of SCEs is a definitive diagnostic test in Bloom syndrome.

Segmental duplication Duplicated DNA block with the same or highly related sequence.

Segregation (Bateson and Saunders 1902) The separation of alleles at a gene locus at meiosis and their distribution to different gametes. Segregation accounts for the 1:1 distribution of allelic genes to different chromosomes.

Segregation analysis A set of different methods to determine the mode of inheritance.

Selection (Darwin 1858) Preferential reproduction or survival of different genotypes under different environmental conditions. Positive selection is in favor of a given genotype as opposed to negative selection.

Selection coefficient Quantitative expression (from 0 to 1) of the disadvantage that a genotype has (compared with a standard genotype) in transmitting genes to the next generation. The selection coefficient (s) is the quantitative expression by which biological fitness (1-s) is decreased; i.e., a selection coefficient of 1 indicates a complete lack of biological fitness.

Selective medium A medium that supports growth of cells in culture containing a particular gene.

Senescence Aging of cells in culture or referring to aging in general.

Sense strand The DNA strand used for transcribing into RNA. It is complementary to the antisense strand. Gene DNA sequences are always written in the 5′ to 3′ direction of the sense strand.

Semiconservative (Meselson and Stahl 1958) The normal type of DNA replication. One DNA strand is completely retained; the other is synthesized completely anew.

Sephardic The Jewish population primarily from Spain, North Africa, parts of France and Spain, and Turkey. They were expelled from Spain in 1492. The name is derived from the Hebrew word Sepharad for Spain (see Ashkenazi).

Sequencing The determination of the sequence of nucleotides in genes or amino acids in protein.

Serpentine A seven-helix transmembrane protein.

Signal sequence The N-terminal amino acid sequence of a secreted protein; required for transport of the protein to the right destination in the cell.

Silencer sequence A eukaryotic DNA sequence that blocks the access of gene activity proteins required for transcription, by forming condensed heterochromatin in that particular area.

SINE (short interspersed nuclear element) Short repetitive DNA sequences (compare with LINE).

SMADS Intracellular proteins that transmit extracellular signals from the transforming growth factor β ligands to the cell nucleus, where they induce downstream TGFβ gene transcription. The name is derived from a protein Sma in *Caenorhabditis elegans* and a *Drosophila* protein Mad (mothers against decapentaplegic).

SNP Single nucleotide polymorphism. A variant in the DNA sequence resulting from a difference in a single nucleotide.

snRNA Small nuclear RNA. One of several small RNAs located in the nucleus; five are components of the spliceosome.

snRNPs (small nuclear ribonucleoprotein particles) Complexes of small nuclear RNA molecules and proteins.

Somatic Refers to cells and tissues of the body, as opposed to germinal (referring to germ cells).

Somatic cell Any cell of an organism that does not undergo meiosis and does not form gametes (as opposed to germ cells).

Somatic cell hybridization Formation of cell hybrids in culture.

Southern blot (Southern 1975) Method of transferring DNA fragments from an agarose gel to a membrane after the fragments have been separated according to size by electrophoresis. (Southern E. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 1975; 98: 503–517).

Speciation (Simpson 1944) Formation of species during evolution. One of the first steps

toward speciation is the establishment of a reproductive barrier against genetic exchange. A frequent mechanism is chromosomal inversion.

Species A natural population with interbreeding of the individuals, which share a common gene pool.

S phase (Howard and Pelc 1953) Phase of DNA synthesis (DNA replication) between the G1 and the G2 phase of the eukaryotic cell cycle.

Spliceosome An aggregation of different molecules that can splice RNA.

Splice junction The sequences at the exon/intron boundaries.

Splicing A step in processing a primary RNA transcript in which introns are excised and exons are joined.

Sporadic Nonfamilial occurrence of phenotype without knowing whether it is genetic or not.

Stem cell An undifferentiated cell able to renew itself by division, retaining the potential for differentiation within a particular developmental pathway. Omnipotent (totipotent) stem cells can differentiate into any cell type; pluripotent stem cells are descendants of totipotent stem cells, which can differentiate into a wide variety of cell types, but not all.

Steroid receptor A transcription factor that responds to a steroid hormone.

Stop codon A codon that terminates translation (UAG, UAA, UGA), originally called nonsense codons.

STS (sequence tagged site) A short segment of DNA of known sequence.

Submetacentric Refers to a chromosome consisting of a short and a long arm because of the position of its centromere (see Metacentric and Acrocentric).

SUMO Small ubiquitin-related modifier. Attaches to other proteins (SUMOylation) to modify their function (Meulmeester E, Melchior F. Cell biology: SUMO. Nature 2008; 452: 709–711).

Superfamily A set of genes or proteins related to each other by evolution.

Synapse A region at the junction of a nerve and a muscle cell or between two nerve cells.

Synapsis (Moore 1895) The pairing of homologous chromosomes during meiotic prophase.

Synaptonemal complex (Moses 1958) Parallel structures associated with chiasmata formation during meiosis, visible under the electron microscope.

Syndrome Within human genetics, a group of clinical and pathological characteristics that are etiologically related, regardless of whether the details of their relationship have yet to be identified.

Synteny (Renwick 1971) Refers to gene loci that are located on the same chromosome, whether or not they are linked.

Tandem duplication Short identical DNA segments adjacent to each other.

TATA box A conserved, noncoding DNA sequence ~25 bp in the 5′ region of most eukaryotic genes. It consists mainly of sequences of the TATAAAA motif. Also known as Hogness box (compare with Pribnow box, in prokaryotes).

T cells T lymphocytes.

Telocentric (Darlington 1939) Refers to chromosomes or chromatids with a terminally located centromere, without a short arm or satellite. They do not occur in man.

Telomerase (Greider and Blackburn 1987) A ribonucleoprotein enzyme that adds nucleotide bases at the telomere.

Telomere (Muller 1938) The terminal areas of both ends of a chromosome containing specific sequences.

Template The molecule that determines the nucleotide sequence for the formation of another, similar (complementary) molecule (see DNA and RNA).

Teratogen (Ballantyne 1894) Chemical or physical agent that leads to disturbances of embryological development and malformations.

Termination codon One of the three triplets signaling the end of translation (UAG, UAA, UGA).

Tetraploid (Nemec 1910) Having a double diploid chromosome complement, i.e., four of each kind of chromosome are present (4n instead of 2n).

Topoisomerase A class of enzymes that unwinds DNA and can control the three-dimensional structure of DNA during replication by cleaving one DNA strand, rotating it about the other, and resealing it (class I) or cleaving and resealing both ends (class II). Used to unwind the DNA double helix at transcription.

Trait An observable phenotype.

Trans-acting A regulatory element acting from a distance.

Transcript An RNA copy of a segment of the DNA of an active gene.

Transcription The synthesis of messenger RNA (mRNA), the first step in relaying the information contained in DNA (see Translation).

Transcription factor DNA-binding proteins that regulate transcription activity.

Transcription unit All of the DNA sequences required to code for a given gene product (operationally corresponding to a gene). Includes promoter and coding and noncoding sequences.

Transduction (Zinder and Lederberg 1952) Transfer of genes from one cell to another (usually bacteria) by special viruses, the bacteriophages.

Transfection Introduction of pure DNA into a living cell (compare with Transformation).

tRNA Transfer RNA, the intermediate between mRNA and protein synthesis carrying a specific amino acid to the growing polypeptide chain in the ribosome.

Transformation A term with several different meanings in biology. In genetics, three main types of transformation are distinguished: (1) malignant transformation, the transition of a normal cell to a malignant state with loss of control of proliferation; (2) genetic transformation (Griffith 1928; Avery et al. 1944), a change of genetic attributes of a cell by transfer of genetic information; and (3) blastic transformation, the reaction of lymphocytes to mitogenic substances (e.g., phytohemagglutinin or specific antigens) leading to cell division.

Transgene A cloned gene introduced into a plant or animal and transmitted to subsequent generations.

Transgenic Refers to an animal or a plant into which a cloned gene has been introduced and stably incorporated. It reveals information about the biological function of the (trans) gene.

Transition A mutation by replacement of a purine with another purine or a pyrimidine with another pyrimidine (see Transversion).

Translation The second step in the relay of genetic information. Here the sequence of triplets in mRNA is translated into a corresponding sequence of amino acids to form a polypeptide as the gene product (see Transcription).

Translocation Transfer of all or part of a chromosome to another chromosome. A translocation is usually reciprocal, leading to an exchange of nonhomologous chromosomal segments. A translocation between two acrocentric chromosomes that lose their short arms and fuse at their centromeres is called a fusion type translocation (Robertsonian translocation).

Transmembrane protein A protein located in the cell membrane with domains inside and outside the cell, usually involved in signal transduction.

Transposition Movement of a genetic sequence, a transposon, from one location in the genome to another.

Transposon A DNA sequence with the ability to move and be inserted at a new location of the genome.

Transversion A mutation with replacement of a purine with a pyrimidine or vice versa (see Transition).

Triplet A sequence of three nucleotides comprising a codon of a nucleic acid and representing the code for an amino acid (triplet code, see Codon).

Trisomy (Blakeslee 1922) An extra chromosome in addition to a homologous pair of chromosomes (see Monosomy).

Tumor suppressor gene A gene involved in control of cell division. It suppresses tumor development when one allele retains its normal function (see Oncogene).

Ubiquitin A small protein that inactivates other proteins.

Ubiquitination Inactivation of a protein by attachment to ubiquitin

Unclassified variant (UV) A variant in the DNA sequence without known functional consequence.

Untranslated region (UTR) Regions at the 5′ or 3′ end of a gene that are not translated into a gene product (5′ UTR, 3′ UTR).

UPD Uniparental disomy. Refers to the presence of two chromosomes derived from the same parent, either as isodisomy when both chromosomes are identical, or heterodisomy when different.

Upstream 5′ direction of a gene (see Downstream).

Variation The differences among related individuals, e.g., parents and offspring, or among individuals in a population.

Variegation Different phenotypes within a tissue.

Vector A molecule that can incorporate and transfer DNA.

Virion A complete extracellular viral unit or particle.

Virus Particle consisting of DNA or RNA of defined size and sequence, enclosed in a protein coat encoded by its genes and able to replicate in a susceptible host cell only.

VNTR Variable number of tandem repeats; a type of DNA polymorphism.

Voltage-gated channel An ion channel that is opened or closed by a gradient of electric current.

Western blot Technique to identify antibodies to proteins, in principle similar to the Southern blot method (q.v).

Wild-type Refers to the genotype or phenotype of an organism found in nature or under standard laboratory conditions, meaning "normal." Not used when referring to humans.

WNT A family of cysteine-rich secreted signal proteins with different roles in cell to cell communication, proliferation, migration, polarity, survival, and self-renewal.

X chromatin (Barr and Bartram 1949) Darkly staining condensation in the interphase cell nucleus representing an inactivated X chromosome (formerly called Barr body or sex chromatin).

Xenogenic Refers to transplantation between individuals of different species.

X-inactivation (Lyon 1961) Inactivation of genes on one of the two X chromosomes in somatic cells of female mammals during the early embryonic period by formation of X chromatin.

X-linked Refers to genes on the X chromosome.

YAC (yeast artificial chromosome) A yeast chromosome into which foreign DNA has been inserted for replication in dividing yeast cells. YACs can incorporate relatively large DNA fragments, up to ~1,000 kb (see BAC).

Y chromatin (F body) (Pearson et al. 1970) The brightly fluorescent long arm of the Y chromosome visible in the interphase nucleus.

Yeast two-hybrid system A technique to identify genes or proteins that interact in function.

Z DNA Alternate conformation of DNA. Unlike normal B DNA, (Watson–Crick model), the helix is left-handed and angled (zigzag, thus Z DNA).

Zinc finger A finger-shaped region found in many DNA-binding regulatory proteins, the "finger" being held together by a centrally placed zinc atom.

Zoo blot A Southern blot containing conserved DNA sequences from related genes of different species. It is taken as evidence that the sequences are coding sequences from a gene.

Zygote (Bateson 1902) The new diploid cell formed by the fusion of the two haploid gametes, an ovum and a spermatozoon, at fertilization. The cell from which the embryo develops.

Zygotene (de Winiwarter 1900) A stage of prophase of meiosis I.

References

Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 6th ed. New York, NY: Garland Science, 2015

King RC, Mulligan PK, Stansfield WD. A Dictionary of Genetics. 8th ed. Oxford: Oxford University Press, 2013

Lodish H, Berk A, Kaiser CA, et al. Molecular Cell Biology. 8th ed. New York, NY: W. H. Freeman; 2016

Passarge E. Color Atlas of Genetics. 4th ed. Stuttgart: Thieme Medical Publishers, 2013.

Rieger R, Michaelis A, Green MM. Glossary of Genetics and Cytogenetics. 5th ed. Heidelberg: Springer-Verlag, 1979

Strachan T, Read AP. Human Molecular Genetics. 4th ed. New York, NY: Garland Science, 2011

Web sites

Genetics Home Reference (http://ghr.nlm.nih.gov/). Accessed July 6, 2017

Glossary of Genetic Terms. Compiled by the Genetics Education Center, University of Kansas Medical Center (http://www.kumc.edu/gec/glossnew.html). Accessed July 6, 2017

Glossaries of genetic terminology WHO (http://www.who.int/genomics/glossary/en/). Accessed July 6, 2017

National Cancer Institute. NCI Dictionary of Genetics Terms. (https://www.cancer.gov/publications/dictionaries/genetics-dictionary). Accessed July 6, 2017

Index

Note: Glossary entries are not included in the index
The entries set in bold are names of diseases.

ABL gene 324
acanthocytosis 334
acardius 130, 131
acetyl coenzyme A 280
N-acetyl-glucosamine 36, 37
N-acetylglucosamine-1-phosphotransferase 288, 290
Acheulian 30
achondroplasia 340
achromatopsia 374, 398
acrocentric chromosomes 144, 158, 159
centric fusion 158
acrocephaly-syndactyly syndrome Typ Saethre–Chotzen 172
actin 334
action potential, cardiac 264
activator-dissociation (Ac/Ds) 222
adaptive immune system 296
adenine (A) 42, 44, 48
deamination 76
adenocarcinoma 316
adenosine deaminase deficiency 42
adenylate 188
adjacent segregation 158
adrenogenital syndrome (congenital adrenal hyperplasia) 368
adrenoleukodystrophy, neonatal 294
agammaglobulinemia, X-linked 310
Alagille syndrome 194, 396
alanine 40, 52
alkaptonuria 393
allele 60, 61, 70
  frequency 122, 124
  allele-specific PCR 60
  codominant 110
  dominant 112
  geographical differences 126
  heterogeneity 388
  population distribution 122
  recessive 112
  allelic exclusion 300, 304
α_1-antichemotrypsin deficiency 266
α_1-antitrypsin 266, 267
  mutations 266
  Pi(Z) allele 266
  synthesis 266, 267
α_1-antitrypsin deficiency (AAT) 266, 267
alternate segregation 158
alternative RNA splicing 168, 169
Alu family (Alu sequence) 210
amino acids 40, 41
amplionic region, Y chromosome 230
anaphase 90, 91, 95, 160
anatomically modern man (AMN) 28
androgen insensitivity syndrome 366
androgen receptor 366
aneuploidy 156, 157
aneusomy by recombination 160, 161
Angelman syndrome 182, 258, 259, 396
ankyrin 334
anosmia 378
antennapedia gene 196, 199
anti-Müllerian 364
antibody diversity generation 300
anticipation 84
anticodon 52
antigen-antibody binding 296, 297
antigen-binding molecules 300
Antley–Bixler syndrome 282, 283, 340
aortic valve disease 194
APC gene 320
Apert syndrome 340
apolipoprotein (Apo) B 168
apoptosis 100, 101
Caenorhabditis elegans 194, 202
arachidonic acid 38
Archaea (archaebacteria) 26
Ardipithecus ramidus 28
arginine 40, 41
array-comparative genomic hybridization (aCGH) 216, 217
Ashkenazi Jewish population 322, 330
asparagine 40, 41
aspartic acid 40, 41
assortive (nonrandom) matings 124
ataxia-telangiectasia (AT) 330, 331
ATM gene 322
ATM protein 80, 330
ATR-16 syndrome 358
auditory system 376
Australopithecinae 28
autonomous replicating sequences (ARS) 137
autonomous transposition 222
autosomal inheritance 114

autosomal trisomies 156, 382
axon 86

B-cell lymphoma 100
B cells 296, 300, 302
backcross, genetic 106
backward slippage 78
transformation 46
transposons 222, 223
bacteriophage genomes 208
bacteriophage φX174 208, 209
baker's yeast *See* Saccharomyces cerevisiae
Balbiani rings 132
band 4.1 protein 334
band-shift assay (electrophoretic mobility shiftassay) 172
Bardet–Biedl syndrome 248, 396
basal cells 378
base analogues 76, 77
basic helix-loop-helix (bHLH) proteins 172,173
BCL-2 gene 202
BCR/ABL fusion protein 324
BCR gene 324
Becker muscular dystrophy (BMD) 338
bioinformatics 206
bivalent 96
Bloom syndrome (BLM) 330, 331
bone development 348
brachydactyly (type 1A) 394
BRCA1/A2 80, 322, 323
breast cancer susceptibility genes 323
brittle bone disease (osteogenesis imperfecta) 346
5-bromodeoxyuridine (5-BrdU) 76
brother/sister mating 128, 129
BTK gene 310
butyrylcholinesterase deficiency 272, 273

C4A/CABcomplement 368
C bands 146
Caenorhabditis elegans 100, 228
apoptosis 202
cellular origins 208
comparative genomics 216
developmental control genes 202, 203
calcitonin gene 168
cAMP response element-binding protein (CREB) 192
cancer
  cell line gene expression patterns 202
  classification 312
  genetic causes 312
  genomes 14, 306, 316
  multistep clonal expansion 312, 313
  origins of 312–316
  cancer genes 312
carbohydrates 36
carcinoma 312
cardiac arrhythmias, inherited 264, 265
caretaker genes 314
caspases 100, 101
cassette model, mating-type switch 88, 89
catfish (Ictalurus punctatus) 378
cdc2 (cell division cycle 2) 140
cDNA 60, 61
  array 224
  cloning 68
  library 68
  preparation 64
  probe 68
  vectors 66
ced-9 gene 202
cell(s)
  components 32, 36
  impaired structure 334–338
  mitosis *See* mitosis
  types 32
cell adhesion molecules (CAMs) 308
cell culture methods 102
cell cycle 90
  control 90
  control systems 90
  mutations 90
cell line dendrogram 214
cell membrane 32
cell membrane-bound receptors 86
cell signaling 86
cell wall 32
cellular signal transduction 186
Centre d'Étude du Polymorphisme Humain (CEPH) 72, 73
centromeres 134
constitutive heterochromatin 134, 135
centromeric DNA 134
centromeric DNA sequences (CEN) 136, 137
CFTR protein 262, 263
channelopathies 264
Charcot–Marie–Tooth disease type 1A (CMT1A) 236
chemosensory epithelia 380
chiasma 94, 132
CHILD syndrome 282, 283
chimeric mouse 176
chloride channel defects 262
chloroplasts 32
cholesterol 38, 39, 280

cholesterol biosynthesis pathway 280, 281
chondrocytes 340
chondrodysplasia punctata type 2 280
CRISPR-Cas 224
chromatid 98
chromatid break 160
chromatin 134
  disorders 238
  DNA packing model 138, 139
  chromatin-associated proteins 228
  reversible changes 180
  topological associated domains 212, 240
chromosome(s) 132
  banding patterns 132, 136
  centromeric sequences 136, 137
  conserved segments 226, 228
  crossing-over 94, 96, 97
  G-banding 146
  genetic disease locations 392
  histone-depleted 134
  interphase 140, 141
  lampbrush 132, 133
  metaphase 92, 93, 95, 134
  mosaicism, Turner syndrome 382
  multiradiate 178
  nonrecombinant 116, 117
  structural aberrations 160
  submetacentric 144, 148
  supernumerary 160
  chromosome-associated proteins 228, 229
  chromosome break 160
  chromosome painting 160
  chromosome puffs 132
chromothripsis 312
chronic myelogenous leukemia (CML) 324
cilia 378, 379
ciliopathies 248
circos plot 316
cis-acting enhancers 184
cleidocranial dysplasia (CCD) 348, 349
clone libraries 66
cDNA 68
coagulation factors 268
cochlea 376, 377
Cockayne syndrome 332
codominant 110
dominant 110
coding strand 48, 52
Coffin-Siris syndrome 238
cohesin 92
cohesin disorders 246
COL1A1 344
COL1A2 344
COL2A1 344, 345
colcemid 150
collagen 344
color vision 370, 374
colorectal cancer 312
common thorn apple (jimsonweed, Datura stramonium) 156
comparative genomic hybridization (CGH) 216, 217
comparative genomics 228
complementary DNA (cDNA) 60, 71
complex diseases
  genome-wide association study 210
  polygenic threshold model 120
condensin 140
cone cells 370, 374
congenital adrenal hyperplasia 368
congenital contractural arachnodactyly 342
congenital flat cornea (cornea plana) 126
congenital hemidysplasia with ichthyosiform erythroderma or nevus and limb defects (CHILD syndrome) 282
congenital intestinal aganglionosis (Hirschsprung disease) 250
connexin 32
Conradi–Hünermann syndrome (X-linked chondrodysplasia punctata type 2) 280, 282
consanguinity 114, 128, 129
consensus sequence 164, 165, 166
constitutive expression 168
contact-dependent signaling 86, 87
copy number variation (CNV) 218, 236
cornea plana (congenital flat cornea) 126
Cornelia-de-Lange syndrome 238
corolla length, Nicotiana longiflora 120
Costal 2 192
Costello syndrome 252
cousins 128, 129
craniosynostosis 340
cri-du-chat syndrome (deletion 5p-) 386
CRISPR-Cas system 224
crossing-over 94, 96, 97
congenital adrenal hyperplasia 368
Crouzon syndrome 340
cultured cells 102
cyclic adenosine monophosphate (cAMP) 188, 189
cyclic guanosine monophosphate (cGMP) 370
cyclin-dependent kinases (Cdks) 90
CYP2 family 274
CYP21 gene 368, 369
CYP21A2 gene 368
CYP pseudogene (CYP21P) 368

cysteine 40, 41
cystic fibrosis (mucoviscidosis) 262, 263
cystic fibrosis transmembrane conduction regulator (CFTR) gene 262, 263
cytochrome P450 enzymes 274
cytolytic T lymphocytes (killer cells) 304
cytoplasm 32
cytosine (C) 42, 44, 48, 58
deamination 76
cytoskeleton 328

Danio rerio (zebrafish) 200
Danisovans 30
Datura stramonium 156, 157
DAX1 gene 364
deafness, congenital 376, 394
deamination 76, 77
debrisoquine-4-hydroxylase (CYP2D) 274
deletion 4p- (Wolf–Hirschhorn syndrome) 386
deletion 5p- (cri-du-chat syndrome) 386
deletion 22q11
(DiGeorge syndrome) 310, 311, 386
delta Δ508 mutation 262
Denisovan hominin 30
deoxyribonucleic acid (DNA) 42
β-D-deoxyribose 42
depurination 76, 77
dermal (intramembranous) ossification 346
desmosterol 282
desmosterolosis 282, 283
deuteranopia 374
diabetes insipidus 396, 397
diabetes mellitus 120, 276
diakinesis 96, 97
diastrophic skeletal dysplasia 126, 127
dicer 174, 175
dideoxynucleotide 62
DiGeorge syndrome (deletion 22q11) 310, 311, 386, 397
7-dihydrocholesterol (7-DHC) 282
dihydrolanosterol 282
dihydrotestosterone (DHT) 364
diplotene 96, 97
disease(s)
  chromosomal locations 392
  X chromosome 392
  Y chromosome 392
  complex *See* complex diseases
  gene identification 220
  geographical differences 126, 127
  susceptibility 323
disomy 156
dispermy 156
disulfide bridge 192, 270
dizygotic twins 130, 131
DMD gene *See* dystrophin gene (DMD)
DMPK gene 254
DNA
  analysis 58–71
  base-excision repair 80
  base pairing 44, 76
  as carrier of genetic information 42
  in chromosomes 152
  cleavage, restriction nucleases 59
  components 44–45
  denaturation 48, 49
  double helix 48
  double strand 45, 48
  fluorescence labeling 62
  genetic information transfer 46
  genetic information transmission 48, 49
  genetic variability 72, 73
  hybridization 66
  intervening sequences *See* introns
  ligated fragments 64
  methylation 178, 179
  mutations 74
  nucleotide chain 44, 48
  polymorphism 70
  processing 80–81
  regulatory protein binding 194
  renaturation 48, 49
  repair 64, 80
  replication 50, 136
  sequencing 64
  structure 48
  transposons 82, 222
  variability 72–79
  zinc finger protein binding 170
DNA-binding domains 172
DNA chips (DNA microarrays) 214
DNA cloning
  cell-based 66, 67
  host cells 67
  vectors 67, 68
DNA excision repair disorders 332
DNA helicases 34, 50
DNA libraries 68, 69
DNA ligase 50
DNA methyltransferases (DNMTs) 178, 238
DNA microarrays (DNA chips) 214
DNA polymerase 64
DNA polymerase III 80
DNA probe 66, 71
DNA-protein interactions 170, 172
DNA replication fork 50, 51

DNAH5 gene 248
DNase I footprinting assay 172
DNMT3B gene mutations 178, 179
dominant 112, 114
dominant negative effect 78
dosage compensation 184
double-stranded RNA (dsRNA) 174
Down syndrome *See* trisomy 21
driver mutations 314
Drosophila melanogaster 132, 156, 196
  bithorax mutation 198
  comparative genomics 216
  embryonic development 198
  embryonic development genes 198
  hedgehog (hh) 192
  larval stages 196
  life cycle 193
  mutant Notch 194
  mutations 196
  pattern-determining genes 198
  polytene chromosomes 132, 133
  segmental organization 196, 197
Duchenne muscular dystrophy (DMD) 337, 339
dynamic genome 222
dyneins 248
dyskeratosis congenita 242
dystroglycans 336
dystrophin 336
  analysis in muscle cells 336, 337
  dystrophin gene (DMD) 336, 337
  deletion distribution 322, 327
  Duchenne muscular dystrophy 337
  germline mosaicism 338
  dystrophin–glycan complex 336, 337

EcoRI 67
EDNRB receptor signaling 250
elliptocytosis 334
elongation, translation 52
embryonic stem cells 176, 177
ENCODE (Encyclopedia of DNA Elements) 212
endochondral ossification 348
endocrine cells 86
endocrine disorders 188
endocrine signaling 86, 87
endolysosome 288
endonuclease 58, 80, 88
endoplasmic reticulum 88, 270
endosome 284
enhancers 172, 173, 178, 180
epidermal growth factor (EGFR) 186, 194
epigenetics 34, 78, 140, 178
epistasis 104
ERM family 328
erythrocyte cytoskeletal proteins 334
Escherichia coli genome 208
etiological (genetic) heterogeneity 220
euchromatin 134
eukaryotes/eukaryotic cells 25, 74
  cell communication 86
  chromosomes 140
  components 32
  DNA replication 48, 50
  gene expression regulation 169, 174, 182
  gene structure 56, 57
  kingdoms 26
  ribosomal RNA 162
  ribosomes 162, 163
  transcript processing 56
  transcription 56, 57, 60
  translation 56
evolution 26
  chromosomes 230
  genes 232
  genomes 232
  hemoglobin 350
  immunoglobulin super family 308
  mitochondrial genome 232
  photoreceptors 370
  pseudogenes 226
  evolutionary medicine 30
  evolutionary strata S1 to S5 230
  evolutionary tree reconstruction 226
ex vivo somatic gene therapy 390
exon 56, 57, 64
  shuffling 226
  trinucleotide repeat expansions 85
external genitalia 364, 365

F_1 h/F2 hybrids 120, 121
factor VIII, IX 268
facultative heterozygotes 114
FADD (Fas-associated death domain) 100
familial adenomatous polyposis (FAP) 320
familial hypercholesterolemia (FH) 284
Fanconi anemia (FA) 322, 330
Fas-associated death domain (FADD) 100
fetal hemoglobin (HbF) 350
FGFR1/R2/R3340
fibrillin-1 gene (FBN1) 342
fibrillin-1 protein (FBN1) 342
fibroblast growth factor (FGF) receptors 186
fibroblast growth factors (FGFs) 186, 340
first cousins 128, 129
fission yeast (Schizosaccharomyces pombe) 90, 91

fluorescence in-situ hybridization (FISH) 153, 155
fluorophores 62, 154
FMR1 gene 231, 256
FMRP protein 256
forward slippage 78, 79
fragile site Xq27.3 256, 257
fragile X mental retardation syndrome 256
fragile X tremor/ataxia syndrome (FXTAS) 256
frameshift mutation 74
Friedreich ataxia 84, 255
fructose metabolism disorders 36
fruit fly *See* Drosophila melanogaster
functional genomics 206
fushi tarazu 196

G_2 phase 90
G-banding (GTG banding) 146
G-protein-coupled receptors 189, 370, 378, 380
galactosemia 36
β-galactosidase 72
gametes 98–99, 109
gametogenesis 98
ganglioside catabolism disorders 38
gap genes 198
gap junction proteins 32
gatekeeper genes 314
Gaussian curve 120
gene(s) 3–4
  amplification, tumor cells 313
  arrangement 374
  control principles 166
  embryonic development 196–203
  evolution 226
  expression 168, 169
  regulation 180, 181
  frequency *See* allele frequency
  function regulation 162–177
  identification 218, 220
  long-range activation by enhancers 168, 169
  mutations *See* mutation(s)
  overexpression 78
  polymorphism 70
  gene knockin 176
  gene knockout 176
  gene locus 106, 120, 126
  quantitative traits 120
gene therapy 224, 390, 391
genetic counseling 110
genetic diagnosis 388–390
genetic drift 124
genetic elements, mobile (jumping genes) 222
genetic (etiological) heterogeneity 314, 388
genetic information, flow of 46
genetic linkage 116, 117
  association analysis 118, 119
  genetic recombination 94
  genetic screen 200
  genetic signal pathways 188–194
  genetic traits, quantitative differences 120
  genetic transmission, atypical patterns 120
genome stability genes 314
genome-wide association study (GWAS) 220, 221
genomic disorders 236, 237
genomic DNA library 68, 69
genomic imprinting 182, 183
genomic instability diseases 330
genomics 206–232
genotype/phenotype relationship 78
germ cells 94, 98
germline mutations 314
  APC gene 320
  breast cancer 322
  cancer 312, 314
  p53 tumor suppressor gene 318
GlcNAc-phosphotransferase (N-acetylglucosamine-1-phosphotransferase) 290
Gleevec 324
globin genes 352
  mutations 350
  unequal crossing-over 350, 351
  alpha-globin genes 352, 353
  beta-globin genes 352, 353
  gamma-globin genes 352, 353
  large deletions 360
  structural changes 356
  tertiary structure 352, 353
  embryonic development 350
glucosamine 36
glucose-6-phosphate dehydrogenase deficiency (G6PD) 126, 273
glucuronic acid 36
glutamic acid 40, 41
glutamine 40, 41
glycine 40, 41
glycogen storage diseases 36
glycolipids 38, 39
N-glycosidic bond 42
Golgi apparatus 195
gonads 203, 364, 365
gout 42
Greenberg skeletal dysplasia 282
growth, intracellular functions 190, 252
growth factors 186
GTG banding (G-banding) 146
GTP (guanosyltriphosphate) 188
GU-AG splicing pathway 56

guanine (G) 43, 45, 55
alkylation 76
guanosine diphosphate (GDP) 188
guanosine triphosphate (GTP) 188
gynogenote 182

Haemophilus influenzae 208
hair cells, ear 376, 377
halophiles 26
“handy man” (Homo habilis) 28
haploinsufficiency 78
haplotype 116
Hardy–Weinberg equilibrium principle 124, 125
HAT medium 102
HBA1/A2 352
hedgehog signaling 192, 193, 282
helix-turn-helix motif 170
helper cell T cells 304
hematopoietic system, gene therapy 390
hemoglobin (Hb) 350
  chain elongation 356, 357
  chains 352
  disorders 350–360
  embryonic 350
  evolution 350
  fetal Hb 350
  fusion 356
  genes 352
  thalassemias 350, 356
  hemoglobin H 273
  hemoglobin Constant Spring 356, 357
  hemoglobin Cranston 356, 357
  hemoglobin Gun Hill 356
  hemoglobinopathies, structural 356
  hemoglobin Saskatoon 356
  hemoglobin Zürich 273, 356
hemophilia A 268, 269
hemophilia B 268
heparan sulfate degradation 290, 291
hereditary deafness 376
hereditary neuropathy with liability to pressure palsies (HNPP) 236
hereditary nonpolyposis colorectal cancer (HNPCC) 78, 320
hereditary peripheral motor-sensoric neuropathy type 1A 236
hereditary persistence of fetal hemoglobin (HPFH) 356, 360
heredity 108
heritability 120
hermaphroditism 366
  Caenorhabditis elegans 202
heterochromatin 134, 180
heterodisomy 258
heterokaryon 102, 332
heterotrimeric G proteins *See* G proteins
heterozygous/heterozygotes 106, 110
  facultative 114
  obligate 112, 114
  parental genotypes 112
  selective advantage 126, 354
hexoses 36, 37
HindII 58
Hirschsprung disease (congenital intestinal aganglionosis) 250, 394, 395
histidine 40, 55
histone acetyltransferases (HATs) 180
histone deacetylases (HDACs) 180
histones 180, 181
HIV virus 304
holoprosencephaly 394, 406
homeobox 172, 198, 199
homeodomain proteins 172
homeostasis, imbalanced 260–274
homeotic selector genes 196, 198
Hominidae 28
hominids 28
hominin 28
Homo erectus 28
Homo floresiensis 28
Homo habilis 28
Homo neanderthalensis 28
Homo sapiens 4, 28
homologous recombination BRCA1 322
  double-strand repair 80, 246
  single-strand breaks 80
homozygous 106, 110
hormone response element (HRE) 170, 172
hormones 86, 186
Hox genes 198
HpaII 178
HTT gene 254
human(s)
  dispersal out of Africa 30, 31
  origins 28
human genome
  architecture 210
  morbid anatomy 392–397
  segmental duplications 210, 236
  sequence types 208, 228
human leukocyte antigen (HLA) 306
Hunchback 198
Hunter syndrome 290, 292, 293, 397
Huntington disease 254, 255, 393
Hurler syndrome 292, 293, 393
hybrid cells 102, 103

hybrid (fused) hemoglobin 356
hydatidiform mole 182, 183
hydrogen, functional groups 44, 45
hydrophilic amino acids 40, 41
3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase) 280
hydroxyl groups 36, 40
hyperglycemia 276
hyperphenylalaninemia 278
hyperuricemia 42
hypohidrotic ectodermal dysplasia, X-linked 184

I-cell disease (mucolipidosis type II alpha/beta) 290, 291
ICF syndrome type 1 178
Ictalurus punctatus (catfish) 378
identity by descent (IBD) 128, 129
immune system 296–311
  cellular response 330
  components 296
  humoral response 296, 297
immunodeficiency diseases, hereditary 297, 298, 310
immunoglobulin(s) (Ig) 296
  binding sites 296
  chains 296, 298
  classes 298
  D-J joining 300
  DNA recognition sequences 302, 303
  duplication events 308
  gene rearrangement 300
  genetic diversity 298, 302
  loci organization 300, 305
  polypeptide encoding genes 298
  protein structure 296, 298
  segments 300
  structure 296, 297, 298
  super family evolution 308
  synthesis, somatic recombination 300
immunoglobulin G (IgG) 298, 299
immunoglobulin M (IgM) 298
imprint erasure 182
imprinting, genomic 182
imprinting center defects 258
imprinting diseases/disorders 182, 258, 259
in vivo somatic gene therapy 390
inbreeding 128
inbreeding coefficient 128
independent assortment 94, 116
independent distribution 108
index patient 110
Indian corn 222
induced mutagenesis, zebrafish 200, 201
induced pluripotent stem cells (iPS cells) 390
inheritance
  autosomal dominant 114
  autosomal recessive 114
  X-chromosomal 114
  Y-chromosmal 114
  mitochondria 232
  monogenic 114
  multifactorial 120
initiation factors 52
innate immune system 296
insertion 74, 79
insertion sequences (IS) 82, 83
insulin 86, 276, 277
insulin-dependent diabetes mellitus 277
insulin receptor 276, 277
insulin resistance syndromes 276
intermediate filaments 334
interphase 92, 134
  chromosome territories 140, 141
  fluorescence *in-situ* hybridization 153, 155
  meiosis I 98
interspersed repetitive DNA 210, 211
interstitial deletion 160
intraflagellar transport (IFT) proteins 248
intramembranous (dermal) ossification 348
introns 56, 57, 84, 85
  GU-AG splicing pathway 56, 57
  trinucleotide repeat expansions 84
inversion 160, 161
  factor VIII 268, 269
ion channels, genetic defects 264
isochromosome 160, 161
isodisomy 258
isoleucine 40
isoniazid 273

Jacobson's organ (vomeronasal organ) 380
Jervell and Lange–Nielsen syndrome 264
jimsonweed (Datura stramonium) 156, 157
jumping genes (mobile genetic elements) 222

Kallmann syndrome 397
Kartagener syndrome 248
karyogram 144, 145, 150
karyotype 148, 151
karyotyping, spectral 154, 155
KCNH2 gene mutations 264
α-keto acid dehydrogenase deficiency 40
ketones 36
killer cells (cytolytic T lymphocytes) 304
kinetochore 92
Klinefelter syndrome (47,XXY) 362, 363, 384

Knirps mutation in Drosophila 198
knockout animal 176, 318
KRAS gene 252
Krüppel mutation in Drosophila 198

lactose 36, 37
lagging strand 50, 142
laminopathies 244
laminins 336
lampbrush chromosomes 132, 133
Langer–Giedion syndrome 394
lanosterol 280, 281, 282
lariat 56
lathosterolosis 282, 283
LDL receptor 284, 285
homology 285, 285
LDL receptor-mediated endocytosis 284, 285
LDL receptor-related protein (LRP) 284
leading strand 50
LEOPARD syndrome 252
leptotene 96, 97
leucine 40, 136
leucine zipper proteins 172, 173
leukemia 324
Li-Fraumeni syndrome 318, 319
ligands 194
light cascade 370, 371
limb girdle muscular dystrophies 336
linkage analysis 118
  LOD score 118, 216
linkage disequilibrium (LD) 118, 216, 278, 358
linker DNA 138
linoleic acid 38
lipid aggregates 38, 39
lipids 38, 39
locus heterogeneity 388
LOD score 118
Loeys–Dietz syndrome (LDS) 342, 343
long interspersed repeat sequences (LINEs) 210
long QT (LQT) syndromes 264, 265
loss-of-function mutation 314
loss of heterozygosity (LOH) 314
  retinoblastoma 326
  tumor suppressor genes 314
lymphocyte culture 150
lymphocytes 296, 297, 300, 304
lymphoma 312
lysine 40, 41
lysosomes 32, 288, 289, 290
  defects 290
  disorders 288
  enzymes 290
  storage diseases 290

main olfactory bulb (MOB) 380
main olfactory epithelium (MOE) 380
maintenance methyltransferase 178
maize 222, 223
major histocompatibility complex (MHC) 304, 306
  classes 304
  gene loci 300, 305
  structure 304
  T-cell receptor interaction 304, 305
malaria 126, 127, 354
male infertility 230
malignant hyperthermia (MH) 272, 273
Manhattan plot 220
mannose 6-phosphate receptors 288, 289
MAP kinases 252
maple syrup urine disease 40
Marfan syndrome 342, 343
Martin–Bell syndrome 256
mating factor (pheromone) 88, 380
mating type conversion 88
maturity onset diabetes of the young (MODY) 276
MDM2 318
medullary thyroid carcinoma (MTC) 250
megalin 284
meiosis 94–97
  daughter cells 94
  medical relevance 94
  mitosis 92, 94
  nondisjunction 156, 157
  ring chromosome 160, 161
  meiosis I 94, 95
  meiosis II 94, 95
  meiotic drive 104
Mendelian rules 104
Mendelian traits 104–110
  deviations of 104
messenger RNA (mRNA) 48, 54, 56
  genetic code 54, 55
  hybridization 56
metabolic disorders 276–294
metacentric chromosomes 144, 148
metaphase 92
metazoa, phylogeny 26, 27
methionine 40, 41
5-methylcytosine methylcytosine-binding proteins (MeCPs) 180
6-methylguanine 76
mevalonic aciduria 280
microdeletion syndromes 310, 386–387
microduplication syndromes 386
microfilaments 334

microRNAs (miRNAs) 174
microsatellites 72, 73
  instability 78
microtubuli 248
Miller–Dieker syndrome 396
minisatellites 72, 73
mismatch repair 80
missense mutations 74, 286
mitochondria 32
  genes 232, 233
  maternal inheritance 232, 260, 261
  mitochondrial diseases 232, 260, 261
  mitochondrial DNA (mtDNA) 260
  deletions 260, 261
  heavy strand 232
  light strand 232
  mutation heteroplasmy 260, 261
  mutation rate 260
mitogen-activated protein kinases (MAPK) 252
mitosis 92, 93
meiosis 94
mitosis-promoting factor (MPF) 94
mobile genetic elements (jumping genes) 222
molecular cytogenetics 152
monosaccharides 36, 37
monosomy X ***See*** **Turner syndrome** 384, 385
Morquio syndrome 393
mouse
  conserved chromosome segments 228, 229
  karyotype 148, 149
  Sry-transgenic male XX 362, 363
  transgenic 176, 177
  trisomies 156, 157
Mousterian 30
mucolipidosis (I-cell disease) 290, 291
mucopolysaccharide storage diseases
  diagnosis 292
  type I (Hurler/Scheie) 292, 293
  type II (Hunter) 292, 293
mucoviscidosis (cystic fibrosis) 262, 263
Muenke syndrome 340, 341
Müllerian ducts 364
Müllerian inhibition factor (anti-Müllerian hormone) 364
multicolor FISH (M-FISH) 154, 155
multiple endocrine neoplasia (MEN) 250, 251
multipoint linkage analysis 216
muscular dystrophies 336, 338
mutator phenotype 316
N-*myc* gene 312
Mycoplasma genitalium genome 208
myoglobin 352
myotonic dystrophy 254, 255

N-*myc* gene (NMYC) 312
Neanderthals 30, 31
neonatal adrenoleukodystrophy 294
nephrosis, congenital 126, 127
neural crest 250
neuroblastoma 312
neurocristopathies 250, 251
neurofibromatosis 1 328, 329
neurofibromin-1/2 252, 328, 329
neuron 86
neuropathies 236
neurotransmitter 86
next-generation sequencing (NGS) 64
NF1/NF2 genes 328, 329
NF-κB (nuclear factor kappa B) 192
Nicolaides-Baraitser syndrome 238
Nicotiana longiflora, corolla length 120, 121
night blindness 372
Nijmegen breakage syndrome 330
no isthmus midbrain mutation 200, 201
non-homologous end joining (NHE) 35, 80
non-insulin-dependent (type 2) diabetes mellitus 276, 277
nonallelic homologous recombination (NAHR) 236, 237
nonautonomous transposition 222
nondisjunction 94, 156, 157, 382
nonpenetrance 114, 326
Noonan syndrome 252, 328
Notch/Delta signaling pathway 194, 195
Notch receptors 194, 195
nuclear factor kappa B (NF-κB) 192
nucleic acids 42, 43
nucleolus 162, 163
nucleosides 42, 43
nucleosomes 138, 139
nucleotide bases 44, 45
nucleotide expansion disorders 254
nucleotides 42, 43
nucleus 32
nullisomy 156

odorant receptors (ORs) 378, 380
Okazaki fragments 50, 51
olfactory nerve cells 378, 379
olfactory receptor protein 378, 379
oligonucleotide probes 68
oncogenes 314
oncogenic translocation 324
oocytes 98

oogenesis 98, 99
oogonia 98
open reading frame (ORF) 54
organelles 32
ornithine transcarbamoylase (OTC) 278
ortholog 226, 227 (*See* paralog)
ossification 348
osteoblast-specific transcription factors (OSFs) 348
osteoblasts 348
osteoclasts 348
osteogenesis imperfecta 346, 347
OTC genes 278
otosclerosis 376
ovarian cancer susceptibility genes 322
ovarian teratoma 182, 183
ovary 364

p21 315
p53 protein 90, 318, 319
p53 tumor suppressor gene (TP53) 318, 319
pachytene 96, 97
PAH gene 278
pair-rule genes 196
palindromic cleavage pattern 58
palmitic acid 38
papain 298
paracentric inversion 160, 161
paracrine signaling 86, 87
parallel DNA sequencing 64, 65
paralog 226, 227 (*See* ortholog)
parental conflict hypothesis 182
parental consanguinity 114, 128
parental genotypes 112
passenger mutations 316
patched 192
pBR322 66
plasmid vector 66
pea plant (Pisum sativum) 104, 105
pedigree analysis 110, 112
  consecutive generations 114
pentoses 36, 37
pepsin 298
peptide bonds 52, 53
pericentric inversion 152, 153
peripheral neuropathy 236, 237
peroxisome 32, 294, 295
  biogenesis 294
  disorders 294, 295
  respiratory chain 294
PEX gene mutations 294
Pfeiffer syndrome 340, 341
phage φX174 208, 209
pharmacogenetics/pharmacogenomics 272
phenylalanine 40, 41, 54
phenylalanine hydroxylase (PAH) 278
phenylketonuria (PKU) 40, 278
pheromone 88, 380
Philadelphia chromosome 324, 325
phosphate compounds 38, 39
phosphodiester 42
phospholipids 32, 38, 39
photo excitation 370, 371
photopigments 374
photoreceptors 370, 374, 375
phylogenetic tree 26–27, 226
pigmentary retinal degeneration 372
Pisum sativum 104
plasma membrane 32, 33
plasmid vectors 66, 174
platelet-derived growth factor receptors 186
pluripotent stem cells 390
PMP22 protein 236
polar body 98
polonies 64
polygenic inheritance 120
polyglutamine disorders 84
polymerase-binding site 164, 165
polymerase chain reaction (PCR) 60, 61
polyposis coli 320, 321
polytene chromosomes 132, 133
population genetics 122
position effect 346
positional cloning 262, 328
posttranscriptional gene silencing 174, 175
postzygotic event 382
Prader–Willi syndrome (PWS) 258, 396
prelingual deafness 376
premodern man 28
prepro-von Willebrand factor 270, 271
preproinsulin 276, 277
prezygotic event 382
primary ciliary dyskinesia 248, 249
primary immunodeficiency diseases 310
primordial germ cells 98
procollagen 344, 345
programmed cell death *See* apoptosis 5, 100, 202
proinsulin 276, 277
prokaryotes/prokaryotic cells 52
promoter 164
proline 40, 41
prometaphase 144
prophase 92, 93, 94, 95, 96
proposita/propositus 110
protanopia 374

protease inhibitors 266, 267
proteome 162
proteome library 68
proto-oncogenes 314
pseudogenes 82
pseudoautosomal regions X and Y 232
pseudohermaphroditism 366
psychrophiles 26
PTPN11 252
Punnett square 108
purine nucleotide bases 42, 43, 44, 45
pyrimidine nucleotide bases 42, 43, 44, 45
pyropoikilocytosis 334

quantitative genetics 120
quantitative trait locus (QTL) 120

RAD51 80, 322
radiation hybrids 102, 103
RAG1/RAG2 302
Ras genes 314
RAS-MAPK signaling cascade 252, 253
RAS mutations 252, 320
RASopathies 252
RB1 (retinoblastoma) gene 326, 327
receptor-mediated endocytosis 288, 289
receptor tyrosine kinase (RTK) 186, 188, 189, 252
reciprocal translocation 152, 158
recombinant DNA technology 66, 68
recombination 80, 94, 96, 116
recombination-activating genes 302
recombination fraction 118
RECQL3 gene 330
red cell membrane disorders, inherited 334
red–green color vision defects, X-linked 374
repetitive DNA 72
replication errors 50, 78
replication (polymerase) slippage 78, 79
replicative transposition 82, 83
resolvase 82
restriction endonucleases (enzymes) 58, 256
restriction fragment length polymorphism (RFLP) 70, 71
restriction map 58, 59
restriction sequence (restriction site) 58, 59
RET protooncogene 250
RET receptor 250, 251
retinitis pigmentosa 372, 373
retinoblastoma 326, 327
retropseudogenes 82
retrotransposition 82, 83
retroviruses 82
Rett syndrome 180
reverse PCR (RT-PCR) 60, 61
reverse transcriptase 60, 82
rheumatoid arthritis 220
RHO gene 372
rhodamine 218
rhodopsin 370, 371
ribonucleic acid (RNA) *See* RNA
ribonucleosides 42
β-d-ribose 42
ribosomal RNA (rRNA) 52, 162
45s ribosomal RNA precursor molecules 163
ribosomes 162, 163
RNA-induced silencing complex (RISC) 174, 175
RNA interference (RNAI) 174, 175
RNA polymerase I/II/III 162, 164, 166
RNase III helicase 174
Roberts syndrome 92
Robertsonian translocation 158, 159
rod cells 370, 371
Romano–Ward syndrome 264
Rubinstein–Taybi syndrome 396, 418
RUNX2 gene 348, 349
ryanodine receptor mutations 272

Saccharomyces cerevisiae 88, 136
Saethre–Choten syndrome 172
Sahelanthropus tchadenis 28
Sanfilippo syndrome (mucopolysaccharidosis type III) 395
sarcoma 312
satellite 144
α-satellites 134, 137
saturated fatty acids 38, 39
Scheie syndrome (mucopolysaccharidosis type IS) 292
Schizosaccharomyces pombe (fission yeast) 90, 91
schwannomin (neurofibromin-2) 328
SCN5A gene 264
sedimentation coefficient, ribosomes 162
segment polarity genes 196, 198
segmental duplications (SDs) 210, 211
segregation analysis 112
single-stand breaks 80
selenocysteine 40, 41
semiconservative replication 50
sensory olfactory nerve cells 378
sensory perception 370–380
serine 40, 41
serpentine receptors 188
Serrate ligands 194
severe combined immune deficiency 310, 311

sex determination/ differentiation 362–369
sex reversal 362
sexual development disorders 366
short interfering RNA (siRNA) 174, 175
short interspersed repeat sequences (SINEs) 210, 211
short tandem repeats (STRs) 72
SHP2 252
Shprintzen syndrome 310, 386
Shprintzen–Goldberg syndrome 342
sickle cell anemia 126, 354, 355, 360
signal transduction 86, 87
Sillence classification, osteogenesis imperfecta 346
single nucleotide polymorphism (SNP 72, 73
sister chromatids 92
situs inversus 248
skeletal dysplasias 340
skin fibroblast culture 102, 103
Slimb 192
Sly syndrome (type VII mucopolysaccharidosis) 290
Smad family 190
small nuclear ribonucleoprotein particles (snRNP) 162
small nuclear RNA (snRNA) 56, 162
small nucleolar RNA molecules (snoRNA) 162
small RNAs 162
Smith–Lemli–Opitz syndrome 280, 283, 394
Smith–Magenis syndrome 236
smoothened 192
sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis 334
somatic cell genetics 102
somatic recombination 300
Sotos syndrome 236
Southern blot hybridization 70–71
SOX9 240
spectral karyotyping (SKY) 154, 155
spectrins 334, 335
spermatids 98
spermatocytes 98
spermatogenesis 98, 99
spermatogonia 98
spermatozoa 98
spherocytes/spherocytosis 335
sphingolipids 38
sphingomyelin 38
sphingosine 38
splice donor site 56
spliceosome 56
splicing 56, 162
spondylocostal dysplasia 194
sporulation 88
SPRED1 252
squalene 280, 281
Sry 362, 363
SRY gene 362, 363, 364
Sry-transgenic male XX mouse 362, 363
stability genes 314
stem cell(s) 390, 391
stem cell therapy 390, 391
stereocilia 376
steroids 38, 39
stomatocytes/stomatocytosis 335
submetacentric chromosomes 144, 148
subtelomeric sequences 136
sucrose 36, 37
sugar derivatives 36, 37
sulfothermophiles 26
sulfur groups 34, 35
supernumerary chromosomes 160
synaptic signaling 86, 87
synaptonemal complex 96, 97
synteny 116

T cell(s) 296, 297, 302, 303
  antigen recognition 304, 305
  differentiation 310
  immune deficiency diseases 310
T-cell receptor 302, 303
targeted gene disruption 176
taste receptor gene family 380, 381
taste receptors, mammalian 380
TATA box 164 165
“Taung child” 28
telomerase 142
telomere 136, 142, 143, 146
telomeric repeat-binding factor (TRF) 242
telomeric sequences (TEL) 136, 137, 143, 160
telophase 92, 134
terminal deletion 160
testes 364
testicular feminization 366, 367
testis-determining factor (TDF) 364
testosterone 364, 366
transforming growth factor β receptor 190, 191
thalassemia 350, 356
thanatophoric dysplasia 340, 341
thermophiles 26
thoracopagus 130, 131
threonine 40, 41
thymine (T) 42, 44, 54
thymosin 1 308
Toll receptor 198
topoisomerase I 50

topological associated domains in chromatin 212, 238
Torso 198
totipotent stem cells 390
trans-acting enhancers 168
transcription 48, 49, 52, 53, 164, 165
transcription-coupled repair 80
transcription factors 166, 172
  yeast cells 90
transcription unit 164, 165
transcriptome 206
transducin 370
transfer RNA (tRNA) 52, 53
transduction pathway 186, 187
transforming growth factor β (TGFβ) super-family 190
transforming principle 46, 47
transgenic mouse 176, 177
transition (compare transversion) 74
translation 52, 53, 74, 75
translocation 158
  fluorescence in-situ hybridization 153, 155
tumor cells 314
transposition 82, 222
transposons (compare transversion) 82, 210
transversion 74
trichothiodystrophy (TTD) 332, 333
triglycerides 38
trinucleotide diseases 84
trinucleotide expansion disorders 84, 256
trinucleotides 84
triplet codon 54
triplets 84
triploidy 156, 157, 182, 183, 384, 385
trisomies 94, 156, 157, 382
tritanopia 374
tropocollagen molecules 344, 345
tryptophan 40
tryptophan synthetase 74
tumor cells
  comparative genomic hybridization 218
tumor necrosis factor receptors (TNFRs) 192
  signal transduction pathways 186, 190
tumor suppressor genes 314, 315
Turner syndrome 362, 384
two-hit hypothesis 326
tyrosine 40, 41
tyrosine inhibitors 324

UBE3A 258
uracil (U) 42, 44, 76
urea cycle disorders 278
Usher syndromes 372

v-erbB 186
variable number of tandem repeats (VNTR, minisatellites) 72, 276
velocardial facial syndrome (Shprintzen syndrome) 386
viral vectors 390
vision, color 370, 371
visual pigments 374, 375
V(D)J recombinase 302
vomeronasal organ (VOM, Jacobson's organ) 380
von Recklinghausen disease 328
von Willebrand disease 270, 271

Waardenburg syndrome 170
whole-genome scan 216, 217
Williams–Beuren syndrome (deletion 7q11.23) 386, 387
wingless (Wnt)/β-catenin signal transduction pathway 190, 191
Wolffian ducts 364
Wolf–Hirschhorn syndrome (deletion 4p-) 386, 387

X-added region (XAR) 230
X chromatin 184, 185
X-chromosomal inheritance 114, 115
  evolutionary strata 184, 185, 231
  gene loci of genetic diseases 372
  genomic structure 230, 231
  homologies 230, 231
  inactivation 184, 185
  mosaic pattern of expression 184, 185
  X-conserved region (XCR) 230
  X-degenerate 230
  X inactivation center 184
  X-inactivation profile 184, 185
  X-transposed 230
xeroderma pigmentosum 332, 393
XIST (X-inactivation specific transcript) 184
XP cells 332
XX males 366, 367
XY female gonadal dysgenesis 366
XY females 366, 367
XYY 384, 385

Y-chromosomal inheritance 112
  genomic structure 230, 231
  homologies 230, 231
yeast cells 90, 91, 136
  cell cycle control 90
  cell division cycle models 90, 91
  cell-type specificity regulation 88
  diploid 88, 89